UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO DE JOINVILLE
COURSE OF MECHATRONIC ENGINEERING

LEONARDO AFONSO FERREIRA BORTONI

CHARACTERIZATION OF A SMART INCLINOMETER SENSOR THROUGH A DEVELOPED HYBRID DOE/MACHINE LEARNING APPROACH

Joinville
2021

LEONARDO AFONSO FERREIRA BORTONI

CHARACTERIZATION OF A SMART INCLINOMETER SENSOR THROUGH A DEVELOPED HYBRID DOE/MACHINE LEARNING APPROACH

This graduation thesis is presented to fulfill the partial requirement to obtain the title of bachelor in the course of Mechatronic Engineering at the Joinville Technology Center from the Federal University of Santa Catarina.

Mentor: Prof. Dr. Mauricio De Campos Porath

Joinville
2021

Dedico este trabalho ao meu pai, minha mãe e meu irmão.

ACKNOWLEDGEMENTS

I would like to acknowledge the Werkzeugmaschinenlabor for providing me the opportunity of working, funding my studies, and encouraging me to develop new approaches and technologies during my period with them. I also would like to thank especially my mentor in Brazil, Professor Maurício, for enabling my first contact with the institute, and thank my supervisor of work, Philipp Dahlem, who always have supported and encouraged my researches and became a friend for life. I thank my friends to whom I've bounded in Aachen, for making my experience abroad much more memorable. I thank my friends from Caverna, who made my life in the university remarkable. I thank my friends from Bolah, friends to whom I share my heart, breath and blood. And I thank my family, on whom my whole being settles.

Aprender dói.

- Meu professor de cálculo II.

ABSTRACT

Smart sensors cluster several electronic and software modules into a single device. Such feasibility leverages that different configurations of communication and routine of measurements might be dynamically and wireless carried out, and through the same approach the measurement's performance might be assessed. If a device's routine of measurement can be modified, and such modifications lead to a different quality of measurement, one may assess how the changing of the routine's parameters affects the measurements' performance through the technique of Design of Experiments. However, the mentioned approach just yields an intuition of how could be the performance of the phenomenon under study for unobserved situations and combinations of parameters, since the technique just recognize the patterns of a phenomenon. In order to state a decision of how a phenomenon will react when facing an yet unobserved situation or combination of parameters, machine learning can be employed to, relying on former patterns remarked by the DOE methodology, predict the phenomenon's response for any situation, if provided enough data. In this work a smart inclinometer sensor with five configurable measurement routine's parameters is fully characterised through an hybrid approach, developed by the author, that complements the DOE analyses with machine learning algorithms. A full factorial design consists the characterization achieved by the DOE, and subsequently the machine learning algorithms rely on the former knowledge acquired to predict the quality of the measurements' performance for any combination of parameters that is found inside a closed interval of parameters' values, namely a delimited volume of parameters. The predictions accomplished are attested with real performances, and the accuracy of the full characterization achieved is considered satisfactory.

Keywords: Machine learning. Small dataset. Ensemble regression. Design of Experiments. Quality of measurement's performance.

RESUMO

Sensores inteligentes agrupam vários módulos electrônicos e de software em um único dispositivo. Tal viabilidade possibilita que diferentes configurações de comunicações e rotina de medições possam ser realizadas de forma dinâmica e sem necessidade de fios, e através da mesma abordagem o desempenho da medição pode ser avaliado. Se a rotina de medição de um dispositivo puder ser modificada, e tais modificações conduzirem a diferentes qualidades de medição, pode-se avaliar como a alteração dos parâmetros da rotina de medição afeta o desempenho das medições através da técnica de Projeto de experimentos (DOE). No entanto, a abordagem mencionada apenas produz uma intuição de como poderia ser o desempenho do sensor em estudo para situações não observadas e diferentes combinações de parâmetros, uma vez que a técnica apenas reconhece os padrões de um fenômeno. A fim de afirmar uma decisão de como um fenômeno reagirá ao ser submetido à uma situação ou combinação de parâmetros ainda não observada, técnicas de aprendizado de máquina podem ser utilizadas para, com base em padrões anteriormente observados pela metodologia DOE, prever a resposta do fenômeno em estudo para qualquer situação, se forem fornecidos dados suficientes. Neste trabalho, um sensor inteligente inclinômetro com cinco parâmetros configuráveis de rotina de medição é totalmente caracterizado através de uma abordagem híbrida, desenvolvida pelo autor, que complementa as análises DOE com algoritmos de aprendizado de máquina. Um experimento fatorial completo consiste a caracterização elaborada pela abordagem DOE, e subsequentemente os algoritmos de aprendizado de máquina baseiam-se no conhecimento anteriormente adquirido para prever a qualidade do desempenho das medições para qualquer combinação de parâmetros que se encontre dentro de um intervalo fechado de valores de parâmetros, a saber um volume delimitado de parâmetros. As previsões realizadas são verificadas com desempenhos reais, e a precisão da caracterização completa atingida é considerada satisfatória.

Palavras-chave: Aprendizado de máquina. Dataset pequeno. Aprendizado de conjunto. Projeto de experimentos. Qualidade do desempenho de medição.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

API – Application programming interface

ASSD – Average of segments' standard deviation

BLE – Bluetooth Low Energy

DOE – Design of Experiments

DOF – Degree of Freedom

GP – Gaussian Process

IC – Integrated Circuit

IoT – Internet of Things

KNN – K-Nearest Neighbor

LOO – Leave-One-Out

MAE – Mean absolute error

MEMS – Micro-Electro-Mechanical Systems

ML – Machine Learning

MSE – Mean squared error

RMSE – Root mean squared error

RT – Radius of Tolerance

SVM – Support vector machine

LIST OF SYMBOLS

$Err_{residual}$	Residual error	
Err_{static}	Static error	
$Err_{transient}$	Transient error	
$\mathbf{g}$	Gravity acceleration vector	$[\frac{m}{s^2}]$
g	Gravity acceleration	$[\frac{m}{s^2}]$
C	Capacitance	$[F]$
ϵ_0	Vacuum permittivity	$[\frac{F}{m}]$
ϵ	Permittivity	$[\frac{F}{m}]$
A	Area	$[m^2]$
d	Distance between parallel capacitive plates	$[m]$
K_s	Spring constant	$[\frac{N}{m}]$
a	Acceleration	$[\frac{m}{s^2}]$
a_s	Acceleration felt by the sensing axis	$[\frac{m}{s^2}]$
θ_s	Angle of the sensing axis	$[^\circ]$
I	Controlled variable	
J	Uncontrolled variable	
k	Amount of factors	
n	Amount of levels	
P, Q, R	Controllable variables of a full factorial design	
O	Output response of a full factorial design	
$\mathbf{x}$	Vector of independent variable	
$\mathbf{y}$	Vector of dependent variable	
x_m	m^{th} Variable of the independent variable vector	

y_j	j^{th} Variable of the vector of dependent variable
$\hat{y}$	Machine learning model's predicted output
D	Dataset
N	Size of dataset
m	Amount of elements in vector of independent variable
f	Regression function
$\mathbb{R}$	Set of real numbers
$\overline{x}_m$	Mean value of the m^{th} independent variable
σ_m	Standard deviation of the m^{th} independent variable
x_{mi}	Value of a sample from the m^{th} element from an independent variable located at the i^{th} line of a dataset
$Z(x_{mi})$	Standardization function of a value x_{mi}
$N(x_{mi})$	Normalization function of a value x_{mi}
x_m^{max}	Maximum value of the m^{th} independent variable
x_m^{min}	Minimum value of the m^{th} independent variable
R^2	Coefficient of determination
ψ	Mean function
κ	Covariance function
T	Conjunct of base learners
h_T	T^{th} base learner of the conjunct T
$H(\mathbf{x})$	Final prediction of a ensemble learner for a given independent variable vector **x**
s_{seg}	Standard deviation of the samples from one segment
t_i	i^{th} sample from a segment
M_{seg}	Total amount of samples of a segment
η	Tolerance value of the Radius of Tolerance metric.

χ	Dummy variable that is assigned with the judgment of the condition stated by the comparison between the predicted and true value regarding the η tolerance.

SUMMARY

1 INTRODUCTION

The world faces nowadays the placement and settlement of the Industry 4.0. Never in the history of science the fastness and progress of technology reached the contemporary standard. Yet, though new and sophisticated technologies are enhanced and developed every day, their deployment as a common and general concept in industry and laboratory usually takes a long time to be settled and to replace outdated and obsolete techniques. The purpose of this work is to conceptualize a modern and efficient approach to recognize patterns, find sweet spots and characterize a process or device.

Aspiring to combine the ascending concepts of Industry 4.0 and machine learning techniques, the team of Large-Scale Metrology from the Chair of Production Metrology and Quality Management, Laboratory for Machine Tools and Production Engineering (WZL) at RWTH Aachen University, developed a hybrid model approach in order to predict the volumetric error of a machine tool (DAHLEM et al., 2020). The idea is to map, with smart sensors, the machine's temperature gradient and its structural thermal displacement response. The data collected from the sensors would train a machine learning model (black box) in order to predict the transient error ($Err_{transient}$) of a rigid body model (white box), aiming to complement an analysis of a movement's static error (Err_{static}) and assess thereby the residual error ($Err_{residual}$) of a machine tool's performance. In order to achieve such accomplishment, the Large-Scale Metrology's team designed its own temperature and tilt's smart sensors. These devices must be capable of providing accurate, sensible and reliable data measurements, since the white and black box models rely on their information to be characterized.

This work was carried out along with the Large-Scale Metrology's team, aiming to provide a full understanding of the two-axis tilt sensor, termed as inclinometer, developed by the laboratory's team. In order to reach a sensible response from the sensor, capable of recognizing thermal displacements, the device measurement parameter's configuration must be exploited until an optimum quality of response is achieved. However, a reliable analysis of the sensor's performance for the driving of all the possible combinations of parameters values is not a trivial endeavour, once that the amount of configurable parameters is numerous and their setting values may range within a wide range.

Inclinometers are devices able to detect spatial rotations of a rigid body. Its applications range from medical scope (DAI et al., 1996) to the assistance into the development of navigation's systems (TANAKA; NISHIFUJI, 1996). The inclinometer's state of the art relies on Micro-Electro-Mechanical Systems (MEMS) technology, that has been proving high sensitivity, reliability and has a affordable cost (ANDREJAŠIČ, 2008). The sensor under study was developed by the Large-Scale Metrology team. It was designed to combine the MEMS measurement technology with the versatility of Bluetooth Low Energy (BLE)

wireless communication protocol, that has its connection stability, power consumption and data transferring rate enhanced with the arising of Industry 4.0. The final device's electronic assembly is composed by a complex of Integrated Circuits (IC) SCA830-D07 from Murata (2020c), responsible to perform the tilt measurement, and the CYBLE-212006-01 BLE microcontroller from Cypress (2019), responsible to broadcast the data received to the final receptor. The connection between these two modules is accomplished through SPI communication protocol.

An immediate assessment of the inclinometer conveyed the knowledge that the sensor developed has a great potential to achieve an accurate and robust performance, nevertheless the numerous measurement parameters' settings must be chosen carefully. Experiments revealed that the intensity of the measurement's noise varies significantly according to the parameters settled. A subtle changing of these parameters may either lead to a severe drop of measurement's accuracy and an unreliable behavior, or may drive to a noiseless, optimum and stable state of the measurement's responses. Nonetheless, the purpose of the existence these several configurations is to, indeed, have a dynamical device, capable of operating and being dynamically configured for different situations. Thus, in order to exploit the maximum potential of the sensor for any objective, a whole characterization, aiming to optimize the inclinometer's performance, must be done in an efficient and smart manner, once that analysing each of the numerous possible parameters' combinations is impracticable.

When diverse factors influence a specific characteristic of an element or process under analysis, Design of Experiments (DOE) is a pretty interesting approach to remark conclusions in regards of how the combination of these several parameters affects the process' quality of response (ANTONY, 2014). The methodology is founded on to systematically change the input variables of the phenomenon under study, and assess how the output performance reacts accordingly. This technique was first structured by Sir Ronald Fischer in the early of 1920, aiming to determinate the effect of different fertilisers on several plots of land (YATES, 1964). Since then, DOE has been widely used in industrial experimentations, targeting enhancements of manufacturing process and finding the cheapest configuration for an optimum solution. Nonetheless, despite the well-established DOE's state of art, such methodology is not the favorite among engineers and managers of the present time, due the fact that the technique requires a lot of time, efforts and is not well suitable for phenomenons with non-linear parameters' interactions.

In order to complement the analyses of the DOE model and exploit its capability of pattern recognition, machine learning techniques can be incorporated into the DOE's evaluation, aiming to predict singularities and behaviors not observed by the DOE approach. Machine Leraning (ML) is, in a generic definition, a collection of methodologies and algorithms that can automatically detect patterns in data and deploy these patterns to predict the future state of a phenomenon under study (MURPHY, 2012). These predictions can be either a decision, a classification or a numerical value. The foundation of ML techniques relies on the deployment of statistical and mathematical models in programming algorithms. These algorithms, when properly applied, can be able to recognize, through the patterns learned, sweet spots of responses. To

achieve a robust and well-defined ML model, capable of describing the patterns of a phenomenon, such algorithms require data that depicts the conditions that had led the phenomenon under study to such a state. The richer in information is this batch of data, termed Dataset, the more significant are the predictions of unknown observations.

This work aims to characterize the response of the inclinometer developed by the team of Large-Scale Metrology, combining DOE and ML approaches in order to develop a hybrid analysis model able to predict the behavior of the sensor for any parameters' combination in a delimited configuration volume. The DOE model provides in this work an assessment of the inclinometer's behavior according to the varying of the device's parameters, so that the approach depicts the influence of each parameter upon the measurement's quality. Besides, the DOE model also provides a dataset with maximum significance for a minimum quantity of data, in order to map and remark the fundamental aspects of the inclinometer's characteristics. Further, this very dataset is employed to train a ML model in order to predict the response for unobserved parameters' configurations, and, thereby, a reliable map of the device's patterns, for a volume of settings delimited by the DOE's theory, is provided. The final goal is to fully characterize the inclinometer's responses through the hybrid DOE/ML approach carried out in this work. The sensor's characterization complies with the Large-Scale Metrology team's objectives providing versatility and knowledge of the inclinometer's behaviour, enabling thus the exploiting of the device to achieve better performances according to the Large-Scale Metrology team's goals and requisites.

To achieve the accomplishment presented, a communication infrastructure between the sensor and a database was implemented. A reliable and systematic methodology to process and evaluate the inclinometer's measurements was developed. Modern techniques to handle with small datasets in machine learning were deployed. The predictions were evaluated with real experiments and their match attests that the ML model based on DOE theory has identified the patterns of the sensor's performances with a satisfactory accuracy.

1.1 GENERAL OBJECTIVE

To map the response of an inclinometer for any measurement parameters' combination into a delimited volume of configuration, deploying a hybrid approach consisted of machine learning techniques founded on design of experiments theory.

1.2 SPECIFIC OBJECTIVES

- To discuss and document the inclinometer device's features;
- To employ design of experiments analysis in order to comprehend the sensor's parameter interactions and remark performance's particularities;
- To deploy machine learning techniques with a small dataset aiming to predict the inclinometer measurement's responses for any parameters' combination in a delimited

volume of configuration.

- To validate the machine learning model founded on design of experiments theory with real experiments in order to attest its reliability and accuracy.

2 THEORETICAL FUNDAMENTS

The inclinometer under study has five configurable measurement parameters, namely interval of measurement, frequency of measurement, number of samples collected at a single measurement, mode of operation and time of measurement request. The noise in the tilt's measurement varies drastically accordingly with the level of each setting. Further in this work is presented evidences that exist interactions among the device's parameters, which increases the complexity of the choice of an optimum setting. *In order to understand the contents discussed in this work, some essential foundations are presented. Is given an introduction of the inclinometer sensing mechanism and in sequence is explained some principles of Bluetooth communication protocol and its performance in smart sensors. Further, one finds a study of design of experiments and the main aspects of its theory and evaluation tools. Concluding, a research of machine learning basic concepts and approached some algorithms, which were employed in the experimental evaluation of this work, is provided.*

2.1 MEMS INCLINOMETER TECHNOLOGY

The tilt sensing mechanism of the device developed by the Large-Scale Metrology team is composed of a MEMS IC inclinometer (Murata, 2020c). The MEMS regards to an instrumentation technology rose at the end of the eighties decade, where mechanical elements were manufactured in a microelectronic scale, in order to compose sensitive elements of microelectronic devices. The foremost distinguishing characteristic of MEMS concerns its structural feature, that almost imitates a mechanical feature. MEMS have channels, membranes, cantilevers, cavity and holes, while an ordinary electronic IC relies on fixed and compressed structures (ANDREJAŠIČ, 2008). The manufacture of MEMS components requires the deployment of micromachining process to perform a sophisticated manipulation of silicon and other substrates, unlike electronic devices, which are fabricated through computer chip technology (PARTNERSHIP, 2002). The interaction with the electronic's micro-world through miniaturization of existing electric-mechanical technology from the macro-world, leads to the possibility of developing new concepts and devices that would not work on the macro-world.

Inclinometers operate through a particular case of the sensing mechanism of an accelerometer. Both the devices work by the very same principle in matters of reading the physical quantity acceleration, but differs from the range of its measurement and from the data processing circuit. The ordinary unity of measurement measured by these devices is gravity g, given as 9.81 m/s^2 (although the real value of gravity varies according to the altitude over the surface of the Earth). Accelerometers are projected to read several different ranges of gravity, for instance from ±0.26 g up to ±12 g (MURATA, 2020a), whereas inclinometers are constrained to the range ±1 g. Thus, since the literature regarding to both the sort of devices is the same, it will

be presented formulations supposing an accelerometer, aiming to idealize a generic case.

Accelerometer MEMS may be studied by four main different means of sensitive transduction: piezoresistive, capacitive, piezoeletric and tunneling. These devices are composed of at least three basic components: a proof of mass, a suspension for the proof of mass and a receptor that correlates the displacement of the mass with an output voltage (AGGARWAL et al., 2010). The accelerometer employed in this work is a capacitive MEMS, SCA830 from Murata (2020c). Especially this capacitive technology is pretty well renowned for its excellent sensitivity and by the fact that its transduction mechanism is by nature insensitive to temperature, as explained by Andrejašič (2008). The author also asserts that the capacitive sensing is independent of the base material, due to the fact that it relies on the variation of capacitance when the geometry of the capacitor is changing.

The basic functionality of a capacitive MEMS accelerometer is presented in the Figure 1. Neglecting the fringing effects of nonuniform distribution near to the edges, the parallel-plate capacitance is given (LYSHEVSKI, 2002):

$$C = \epsilon_0 \epsilon \frac{A}{d} = \epsilon_A \frac{1}{d} \tag{1}$$

Where ϵ is the permittivity, A is the electrode plate's area, d is the displacement between the plates and $\epsilon_A = \epsilon A$. A change of any of these parameters will result in a change of the capacitance. Humidity and chemical sensors may be founded on the variation of ϵ. Accelerometers works based on the change of d or A. As mentioned before, MEMS accelerometers are consisted of a proof of mass with movable capacitive plates. This structure is attached through a mechanical suspension system to a reference frame. The movable plates stand between the fixed outer capacitive plates. Thus, when a movement by the proof of mass is performed, the gap $x1$ and $x2$ of the fixed plates regarding the movable plate is displaced by a distance of value x. This variation results in a change of the capacitances $C1$ and $C2$, as depicted in the Figure 1, and can be equated as:

$$C_1 = \epsilon_A \frac{1}{x_1 + x} \quad , \quad C_2 = \epsilon_A \frac{1}{x_2 - x} \tag{2}$$

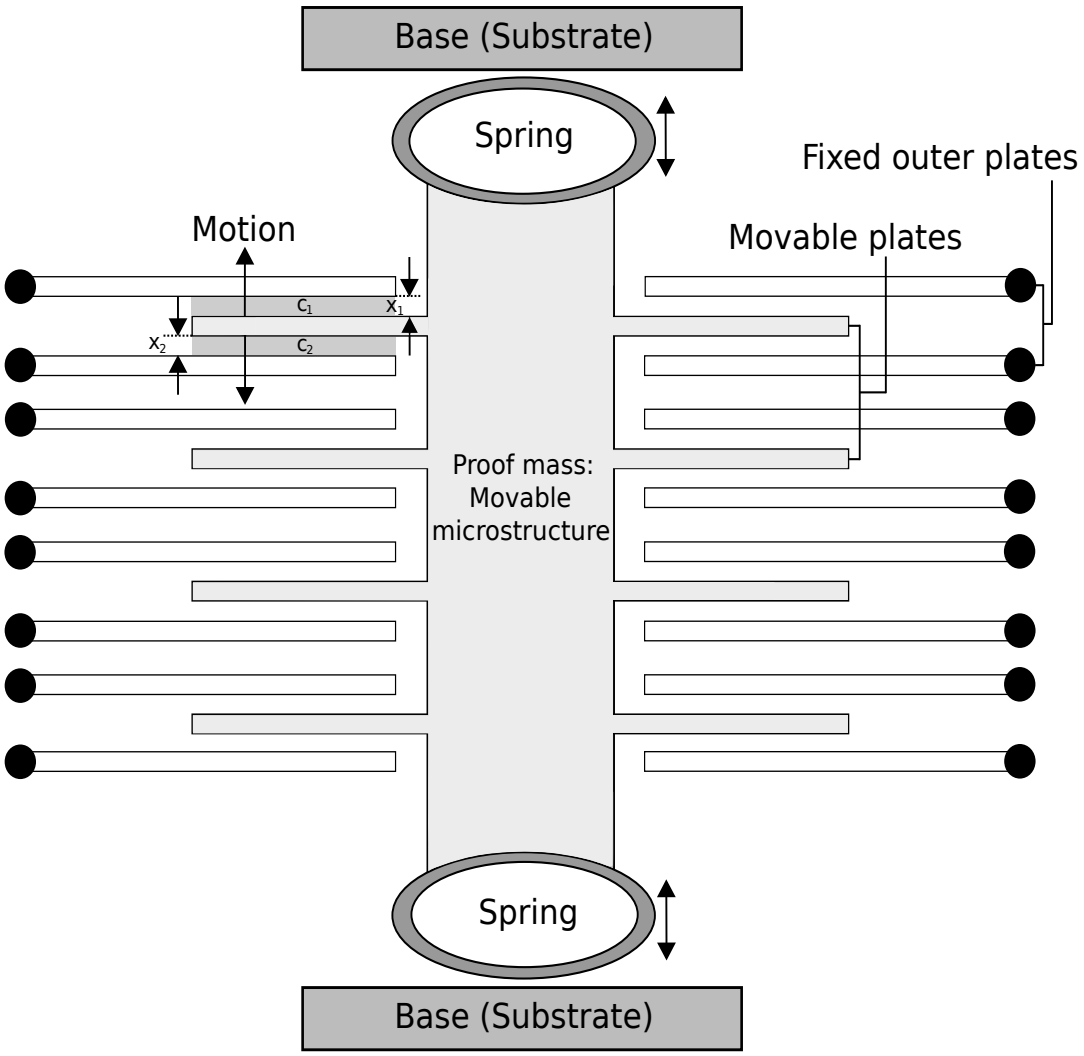
Supposing a perfect alignment of the movable mass: $x1 = x2$. Therefore when there is no acceleration, $x = 0$, the difference of capacitance $C_1 - C_2 = \Delta C$ is zero. When a movement exists, considering $x1 = x2$, the capacitance's difference is found to be:

$$\Delta C = C_1 - C_2 = 2\epsilon_A \frac{x}{x^2 - x_1^2} \tag{3}$$

Measuring ΔC, one finds the displacement x through the solving of the nonlinear algebraic equation:

$$\Delta C x^2 - \epsilon_A x - \Delta C x_1^2 = 0 \tag{4}$$

Figure 1 – MEMS accelerometer structure. The movable proof of mass is attached through springs to a base substrate. The structure can move only up and down. The springs perform the transduction of the changing capacitance C_1 and C_2 between the movable and outer fixed capacitor plates



Source: adapted from Lyshevski (2002, p.122)

This equation can be simplified for small displacements, where the term $\Delta C x^2$ become negligible. Thus, omitting $\Delta C x^2$:

$$x \approx \frac{x_1^2}{\epsilon_A} \Delta C \tag{5}$$

The Eq. 5 express how the capacitance's difference varies in function of a displacement x. For an ideal spring, according to the Hook's law, the module of a restoring force F_s is a function of the displacement x of the spring, so that $F_s = K_s x$, where K_s is the spring constant.

From Newton's second law of motion, neglecting the air friction, this motion of the spring's displacement can be formulated as $ma = md^2x/dt^2 = K_s x$ (LYSHEVSKI, 2002). Therefore, the module of the acceleration as a function of the displacement can be presented as:

$$a = \frac{K_s}{m}x \tag{6}$$

Applying the Eq. 6 on the formulation of the MEMS accelerometer, where m is the mass of the proof mass structure and K_s is the spring constant of the spring which sustains the movable mass structure, the displacement x can be correlated to the module of acceleration as a function of the variety of capacitance between the plates by deploying 5 in 6, resulting:

$$a = \frac{K_s x_1^2}{m\epsilon_A}\Delta C \tag{7}$$

The Eq. 7 depicts the functionality principle of an accelerometer. The difference of capacitance can be read and converted to voltage by several arrangements of electronic circuits, which will not be approached in this work. Since K_s, x_1^2, m and ϵ_A are known values, the relation between the module of the proof mass' acceleration is directly correlated with the capacitive plate's capacitance difference.

As commented before, the output electronic circuit of an accelerometer is charged to interpret the capacitance voltage of the plates and to provide to the user a meaningful value of acceleration through the eq. 7. In inclinometers, this value is given between the range ± 1 g. One can thus relate the acceleration given by the sensing axis of the device with the gravity's acceleration vector of the Earth **g**, in order to define the sensing axis' inclination regarding the gravity vector.

The Figure 2 illustrates the formulation of the conversion from gravity to tilt angles. First, is assumed the tilt sensor laid on the x-axis, in such a manner that the sensing axis points to the negative direction of the x-axis. The Earth's acceleration vector points downward on the y-axis. In the situation of the left side of Figure 2, the sensing axis is not tilted and no acceleration is felt by the device. When there is a slope θ_s of the sensing axis regarding the x-axis as shown in Figure 2 middle, the vector acceleration felt by the MEMS mechanism has a magnitude equal to its projection on the acceleration's gravity vector. The module of the acceleration felt by the sensor a_s is hence a sinusoidal response of the gravity acceleration module g, such as:

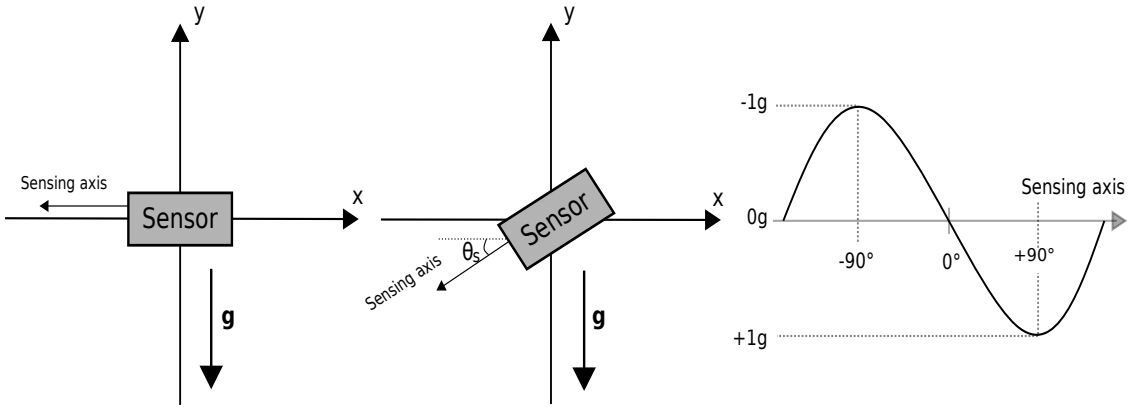
$$a_s = sin(\theta_s) \cdot g \tag{8}$$

Since the MEMS electronic circuit provides the value of a_s as output response, one can find the inclination angle of the device regarding the gravity through the inverse sine operation:

$$\theta_s = \arcsin(\frac{a_s}{g}) \tag{9}$$

In this work, the eq. 9 is widely used in the infrastructure created by the author to read the response of the inclinometer.

Figure 2 – At the left side, a representation of a inclinometer under effect of gravity and there is no tilt of its placement regarding the coordinate system. At the center is portrayed the sensor's displacement on the reference system when there is a tilt angle θ_s regarding the x-axis. At the right is a graphic illustration of the device's mathematical response model for all the possible real values of θ_s.



Source: Autor

Concerning the quality of the tilt measurements, an interesting aspect of the MEMS accelerometers is the noise characteristics of these devices, which influences their performances especially under lower g operating conditions. Mohd-Yasin, Korman e Nagel (2003) made an analysis of the noise aspect of MEMS accelerometers and concluded that there are three primary noise sources typical of these sensors. The first, is due to the mechanical vibration of the transductor springs, the second is related to the signal conditioning circuitry and the third is from the measurement system itself. In their work, the authors also detected that the source of sporadic high peaks noise is related with the squared wave oscillators inside the accelerometer conditioning circuitry.

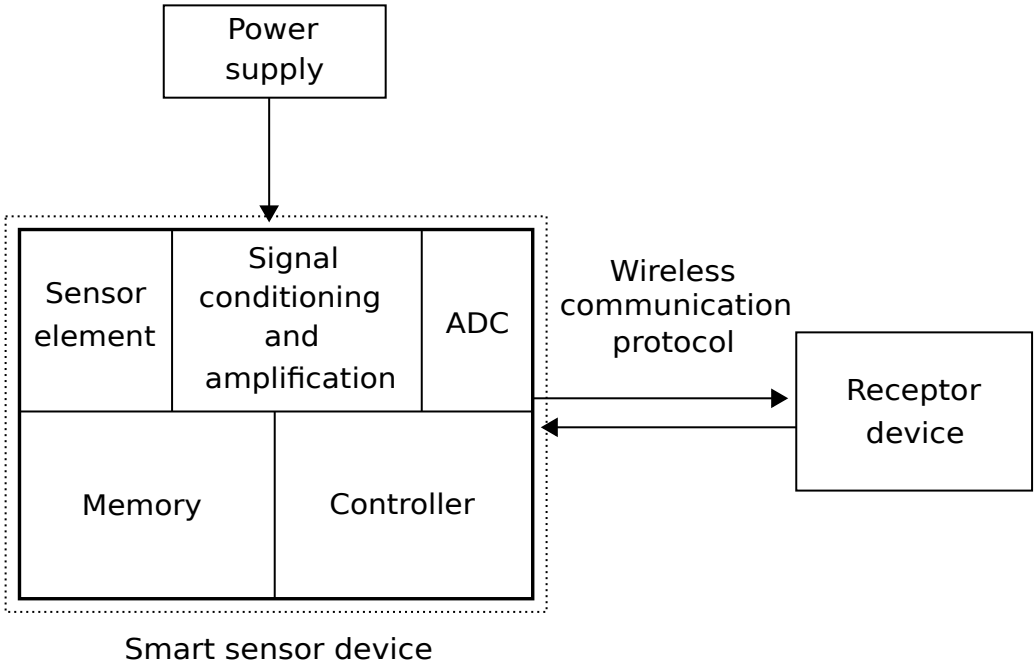
Since these noise characteristics are intrinsic of the MEMS technology, the only easy approach to optimize the inclinometer response is through finding an optimum operating configuration of the device. All the strategies and methodologies developed and presented in this work aims to map the noise response of all the possible configurations of parameters of the device.

2.2 SMART BLUETOOTH LOW ENERGY SENSORS

The inclinometer was projected as a smart sensor. Specially this category of sensors devices was responsible for the consolidation of the industry 4.0 and Internet of Things (IoT). This

accomplishment is due to the robustness and practicality provided through the assembly of several electronic and software modules into one single sensor device, so that complex infrastructures of data exchange have become accessible and of easy management. While ordinary sensors normally just have the sensor element and a wired communication protocol, smart sensors cluster the sensing element, the amplification and signal conditioning module, an Analog-Digital converter, a memory system, a logic control unity and typically a wireless communication protocol in one single device (FRANK, 2013). Besides these features, some modern smart sensors have improved their power supply module from wired sources to battery, targeting a dynamical and totally wireless device. The Figure 3 depicts a typical assembly of smart sensors' modules.

Figure 3 – The figure portraits the general assemble and interactions of a smart sensor. The elements inside the dashed contour compose a typical smart sensor device. The device is energized through a power supply module as battery or wired mechanism. It exchanges information with the receptor device by means of some wireless communication protocol.



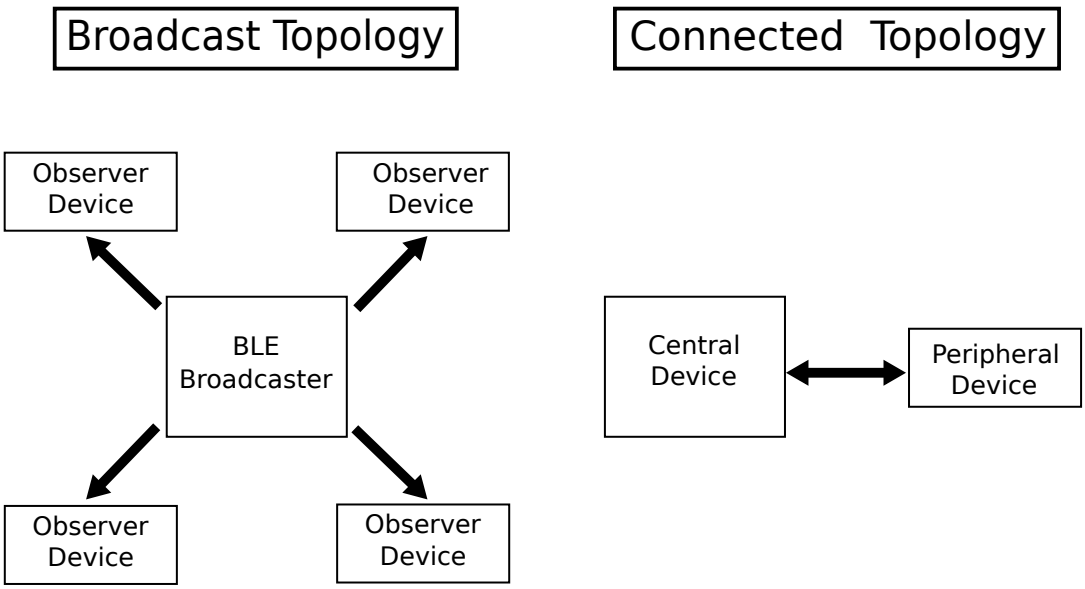
Source: adapted from Frank (2013, p.6)

As commented previously, the sensor developed by the Large-Scale Metrology team has several features that classify it as a smart sensor. The device has a battery power supply mechanism, a programmable logic unity, a Bluetooth Low Energy communication protocol that broadcasts the processed data and a sensing MEMS capacitive mechanism that provides digital data through SPI communication protocol to the logic unity. By exploiting these features, an entire infrastructure of communication could be developed to read and program the device with

a robust performance.

The inclinometer developed has a Bluetooth Low Energy controller from Cypress (2019). This module is charged to perform the logic processing of the data received from the sensing element, store this information and broadcast it through BLE technology. Bluetooth is a wireless communication protocol that allows devices to exchange information by means of radio links (GUPTA, 2013). It is a global standard communication technology acknowledged for its short range performance, low power consumption, low cost and small form factor. Bluetooth Low Energy is an enhancement of the Bluetooth technology added as a part of Bluetooth 4.0 core specification. Although BLE technology is a branch of the classic Bluetooth, it has an entire different lineage and design goals. As detailed by Gupta (2013), the BLE technology is fully optimized from its core until the application level to ensure the minimum power consumption. This enhancement had led to a complete redesign of several key components, in order to assure the minor requirement of power as possible.

Figure 4 – At the left, a representation of the BLE broadcast mode, that is consisted of the broadcaster device propagating its small packages of data to several devices able to pick it up. At the right, a representation of the BLE connected mode, where the central and peripheral device establish a connection and exchange large packages of data.



Source: adapted from Townsend, Cufí e Davidson (2014, p.9 and p.11)

The BLE protocol communicates through two different methodologies: Broadcasting and Connections. As explained by Townsend, Cufí e Davidson (2014), in Broadcasting mode small packages of data can be sent out in one-way to any scanning device or receiver in the

listening range. This concept implies that the broadcaster device propagates its information to any observer capable of picking it up. On the other hand, the observer device can only read the package of data broadcast. This configuration is the only method in BLE that allows the same single information can be read for several devices at the same time. When there is a need of periodical exchange of larger packages of data, the Connection mode is more suitable. In this configuration, the peripheral device sends periodically advertising packages. These packages are picked up by the central device which scans for it, and, when convenient, initiates a connection. Once the connection is established, the peripheral ceases its propagation of advertising and the two devices begin to exchange data in both directions. The Figure 4 summarizes the behaviour of these BLE communication approaches.

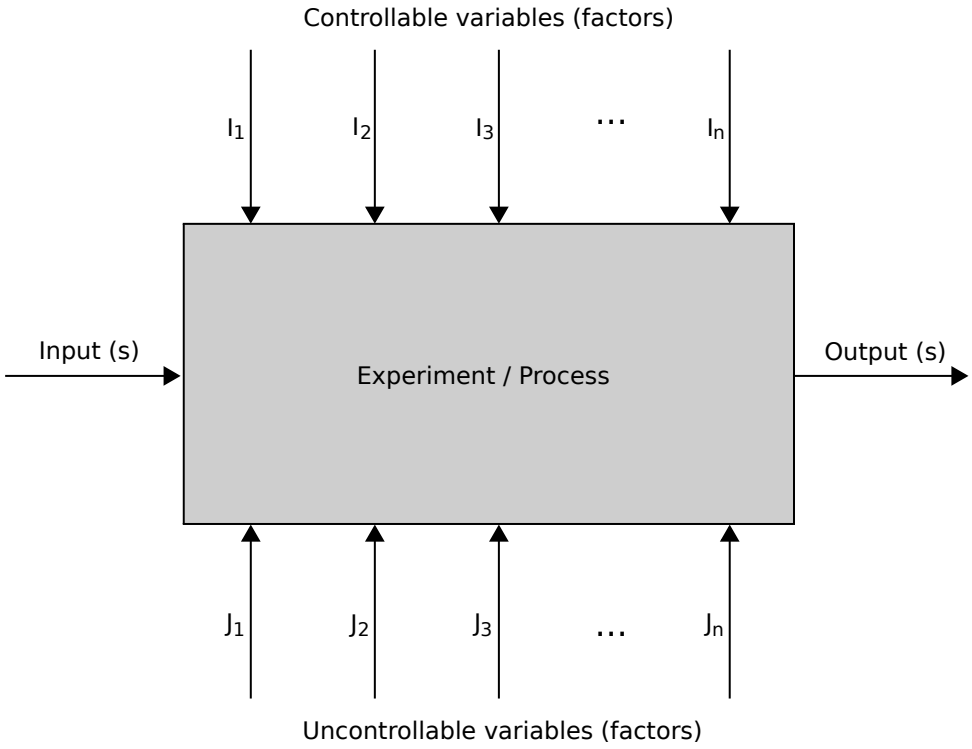
In this work, the connection mode was widely exploited. Periodical routines of reading and reprogramming the inclinometer's configuration were accomplished through the continuous exchange of larger packages, which could last a period of several hours. The BLE technology presented itself high efficiency at the saving of power consumption, so that single units of battery could supply months of experiments.

2.3 DESIGN OF EXPERIMENTS

Design of experiments is a collection of methodologies and approaches developed in order to structure the driving and evaluation of processes and experiments. In the engineering context, experiments are conducted to survey, identify patterns, estimate and confirm assumptions of real models. In the industrial environment, processes are assessed in order to improve a product and quality of a process, and to cheapen the production cost. Whether it is a process or a scientific experiment, DOE provides a systematic procedure to make deliberate changes of the input parameters or variables of a model, termed as factors, aiming to characterise how the output functional response varies accordingly (ANTONY, 2014). Some of the input factors may strongly influence the output response, some may have a moderate impact and some may have no effect at all. Thus, the final goals of DOE methodologies are to comprehend, through empirical experiments, which arrangement of variables most affects the output response, and to determine the levels of these factors that lead to a satisfactory performance.

In real-life conditions, an experiment's performance is characterised by the correspondence between input variables and the output response. The process' variables, also called factors, are classified into two groups: Controlled variables (represented by I) and uncontrolled variables (represented by J). The former regards an easily varying setting of an experiment, as an example the interval of time that a sensor is left collecting some data. The second concerns to hardly or too costly varying setting, as the environment's vibration or temperature. An optimum process or experiment response is achieved when optimal settings of controlled variables that minimise the effect of uncontrolled variables are reached. The Figure 5 illustrates the base structure of an experiment's model.

Figure 5 – Elementary model of an experiment or process. The input arrow represents the settings and arrangements of the experiment. The controllable and uncontrollable variables, represented respectively by I and J, influence the experiment's output response.



Source: adapted from Antony (2014, p.8)

There are some important terminologies whose the classical DOE is founded. As mentioned previously, a factor is a process' variable, either controllable or uncontrollable. A factor may be set to different levels. Here the term 'level' refers to a specific value or degree of the factor under analysis. The changing of the average output response due to a change in the levels of a factor is called 'effect of a factor'. When the effect of one factor at the output response is different for different levels of another factor, it is said that exists 'interaction' between these two factors. In order to understand and map the response of an experiment, it is important to vary the factors simultaneously at their respective different levels, aiming to obtain a reliable and predictable estimation of the factor's effects. The performance of an experiment whose the variation of these several factor's levels composes a unique arrangement of parameter's configuration is termed as 'trial'. A DOE evaluation approach exploits the response of several different combinations of trials.

2.3.1 Replication, blocking and randomization

In order to establish a reliable and systematic methodology, there are three principles of experimental design that shall be followed: Replication, blocking and randomization. The first

two assist the improvement of the precision of an experiment, and the last is employed to reduce experimental bias (DEAN; VOSS; DRAGULJIĆ, 2017).

Is considered as replication the repetition of an entire experiment under more than one condition, so that the output response of interest can be accurately estimated and the associated variability observed. An example of replication can be noted in an illustrative chemical experimentation. It is supposed that one wishes to measure the pH of a material under influence of a developing acid. The experiment consists of applying a sample of the acid on the material, and then measuring the material's pH. The deploying of replication in this procedure could be to apply a sample of the acid on several samples of the material, and then to measure the pH of each of them, so that the experiment is replicated. A common misunderstanding of replication is to, in this example, just repeat the measurement itself of one single material's pH. An important property of replication is that it allows an accurate estimation of the experimental error, a term that designates the difference of what would be observed if the same experimental settings had been applied several times at the same conditions (ANTONY, 2014). Besides, the precision of the output response can be improved through the averaging of each replica's response.

The principle of blocking is related to the concept of arranging similar or homogeneous trials of experiments into groups (or blocks). As explained by Bisgaard (1994), this approach is useful to eliminate extraneous variation due to noise factors and thereby improving the efficiency of the experimental design. The idea consists of grouping a trial with a set of relatively homogeneous experimental conditions. These groups can be sorted per batch of material, operator, supplier, etc. Also, it can be grouped under the same experimental conditions, i.e. same day, same time, same place, etc; so that unwanted sources of variability, such as batch-to-batch or day-to-day, can be noted and pondered. Conceptualizing it in the pH's measurement example of the material under the test of a developing acid, the blocking approach could be set to so arrange the trials of experiments, that each trial were carried out in a different place from the other trials. Another strategy would consist of each trial having the material taken from a batch different from the others trials' batch, so that each trial would experiment the variation, regarding just the material, of an specific batch.

The last principle of experimentation is randomization. This one, is considered the most important rule to avoid experimental and systematic biases (DEAN; VOSS; DRAGULJIĆ, 2017). Randomization consists of randomly assigning of subjects, material, levels and even the trial's blocks execution order. By properly randomising the experiment, possible noisy factors present in the process may be equally distributed among all the levels of a factor, averaging thereby the effect of noisy singularities. Back to the pH test example, it is supposed that there are three different levels of the acid, and the material to be tested is supplied by three different providers. Without randomizing the procedure, the assignment of the acid's application could be made through a systematic order, so that the lower level of the acid would be deployed on the material from the first supplier, the middle level deployed on the material from the second supplier, and the highest level deployed on the material provided by the third supplier. One may observe

that this approach would lead to a systematic bias, once that the materials' manufacturing is not exactly equal among the providers. The effect in the output response due to the setting of different levels would have a systematic response that could not be generalized. The deploying of randomization, in the portrayed situation, would imply to sort the application of different levels of acid among the materials from the different providers in a random sequence. One may note that the noisy singularities of the material would remain intrinsic to its origin, nonetheless its effects would be averaged in all the levels of the experiment.

2.3.2 DOE architecture

Each of the principles of experiments explained in Subsection 2.3.1 composes the DOE's planning phase. The DOE's methodology is founded on four phases: Planning, designing, conducting and analysing. At the planning phase, the output responses are defined and the factors that influence the experiment under study are determined. Hence, an inquiry must be done in order to consider which of the factors can be treated as a controllable variable I, aiming to render known the factors that can be regularized. Afterwards, the procedure to handle with the experiments is architected based on the principles of replication, blocking and randomization. The following phase is to design the experiment itself.

There are three main designs to orchestrate an experiment: Screening, fractional factorial and full factorial. The first two are strategies to analyse a large number of process or design parameters (factors) with a relative small number of experimental runs or trials, and the last, is an approach that exploits all the possibilities and combinations of parameters and its levels. Since this last strategy was deployed in this work, the following explanations will have it as scope. More information about Screening and fractional factorial designs can be found in Antony (2014).

A full factorial designed experiment consists of the running of trials from all the possible combinations of levels from all the factors. It is said that an experiment that has k factors, each one with n-levels, will need k^n trials to carry out a full factorial design (ANTONY, 2014). It is important to note that, although costly, the exploit of all possible parameter's interactions certainly will expose patterns and relationships among the factors, in such a manner that it allows the achievement of an ideal configuration for an optimum output response. A common schematic of full factorial design, deployed to survey an experiment, is a k^2 model. This approach is normally used to study the behaviour of the experiment at the boundaries of the factors' levels, so that the highest levels are denoted as +1 and the lowest levels denoted as -1. This nomenclature facilitates the organization of the experiment's trials in tables. As an example, is supposed an experiment employing a 3^2 full factorial design. The controllable variables are P, Q and R. The output response under interest is O. An arrangement of the experimental table for this design is presented in Table 1.

Once defined and settled the experimentation model design, one may revise how the three principles of experimentation, i.e. replication, blocking and randomization, works with

Table 1 – 3^2 full factorial design example. The terms P, Q and R are controllable variables and O is the output response under analysis.

Trial number	P	Q	R	O
1	+1	+1	+1	x_1
2	+1	+1	-1	x_2
3	+1	-1	+1	x_3
4	+1	-1	-1	x_4
5	-1	+1	+1	x_5
6	-1	+1	-1	x_6
7	-1	-1	+1	x_7
8	-1	-1	-1	x_8

the determined design, in order to profitably arrange them. Subsequently, the conducting phase can be commenced. All the planned experiments are carried out at this phase. There are some suggestions in the literature (ANTONY, 2014) concerning good practices during the performance of a designed experiment, such as: The person responsible for the experiment should be present throughout it; the trials must be always under monitoring; to record the output response on the prepared data sheet or in a computer; and to record and analyse any experimental deviation of an unusual occurrence. When every trial is concluded and the output response treated and computed, one may exploit the obtained results at the analysing phase by means of several graphical approaches.

2.3.3 Evaluation of experiments

The analysing phase consists of the interpretation of the treated data acquired by the experiments. By the means of deployment of statistical methodologies and graphical resources, the patterns of the experiment or process under study may be remarked and thereby the phenomenon can be characterised. An accurate analysis would employ, besides the graphical approach, statistical models in order to assess the variation of the output response according to the changing of levels. Only the graphical approach was employed in this work, since the amount of controllable factors were not numerous and a visual interpretation would grant an easier and satisfactory evaluation. Thus, the following presentations of evaluations do not exceed the scope of graphical methods. More details about statistical methodologies that evaluate experiments can be found in Dean, Voss e Draguljić (2017)

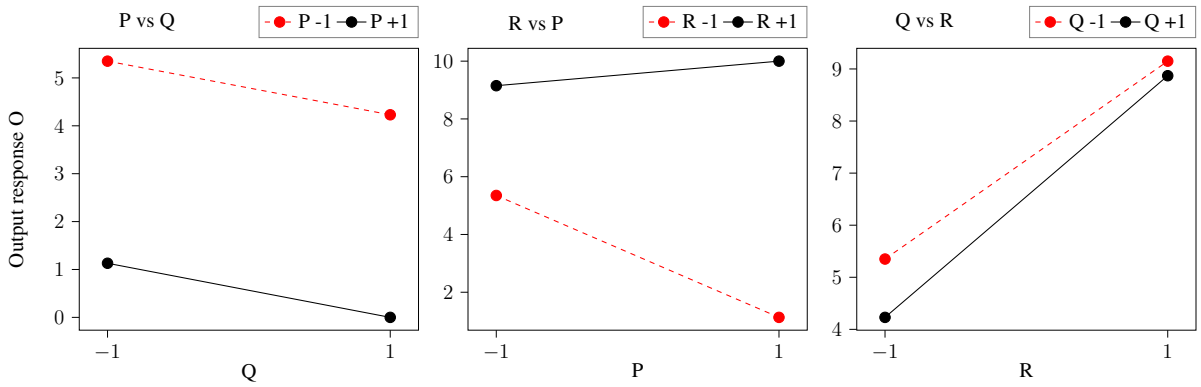
There are several different graphical models that can be employed in order to assist the patterns recognition of an experiment. Antony (2014) presents some of the widest employed models, such as: Main effects plot, interactions plots, cube plots, pareto plot of factor effects etc. In this work were employed interactions and cube plots. The interactions plot is a powerful

graphical tool that plots the output response for all the possible settings combinations of two factors. Supposing a 3^2 full factorial design experiment and its results portrayed at the table 2, with P, Q, R as controllable variables and O as output response, its interactions plot would be as depicted in the Figure 6.

Table 2 – Example of a 3^2 full factorial design experiment with its respectives output results. The terms P, Q and R are controllable variables and O is the output response under analyses.

Trial number	P	Q	R	O
1	+1	+1	+1	9.72
2	+1	+1	-1	0.00
3	+1	-1	+1	10.0
4	+1	-1	-1	1.13
5	-1	+1	+1	8.87
6	-1	+1	-1	4.23
7	-1	-1	+1	9.15
8	-1	-1	-1	5.35

Figure 6 – Interactions plot regarding the designed experiment represented in the Table 2. Here, one may simply visualize the interactions between two factors, when the remaining factors are set in a specific level. In the present case, for each interaction plot, the remnant factor is set at the level -1.



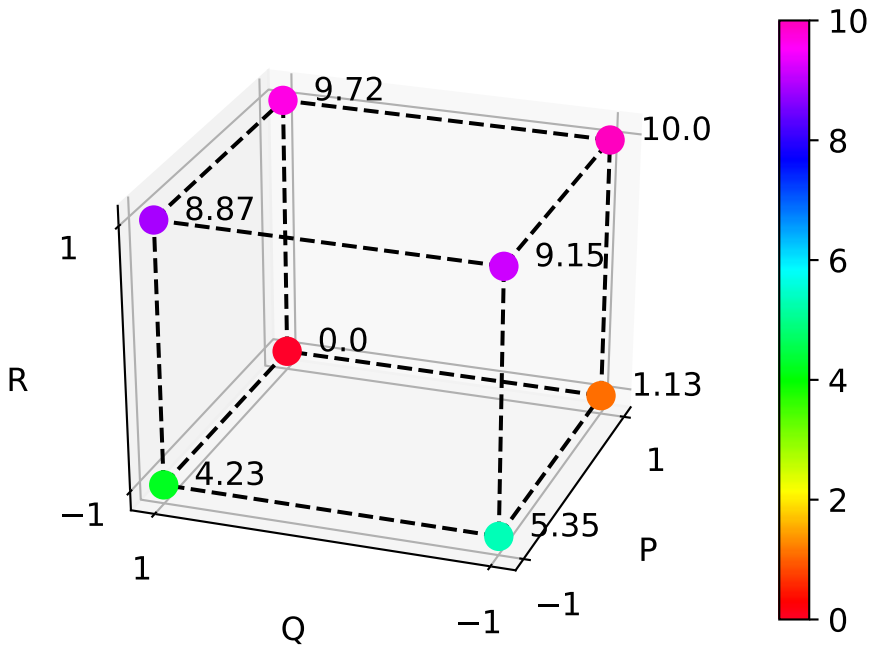
Source: Author.

The interactions plot presented in the Figure 6 is developed in order to determine whether exists or not an interaction between two experiment's parameters. An interaction can be remarked when the change in the output response from the setting low level to high level of a factor does not depend on the level of the other factor. Each interaction plot is so designed that only two factors in the plot vary. The remaining factors must remain set at a unique level, either +1 or -1. In the example of the Figure 6, for each plot, the remnant factor is remained at -1. If the lines of an interaction plot are parallel, it implies that there is no interaction between the parameters plotted. Differently, if the lines are non-parallel, an interaction exists between

the factors. The greater the degree of departure from parallelity, the stronger the interaction effect (ANTONY, 2014). For the interaction P vs Q when R is set in -1, in Figure 6, one may note that the lines are perfectly parallel, and therefore, there is no interaction between P and R. In another hand, for the plot concerning R vs P when Q is set in -1, it is huge the degree of non-parallelism between the lines. Thus, the output response is strongly sensitive for the setting combination of P and R. An interesting effect occurs in the Q vs R plot, when P is set in -1. One may note that although the lines are definitely non-parallel, it is not great the degree of departure from parallelity. This fact implies that there exists interaction between Q and R, but it is minor. The output response will not be too penalized due to an unfavourable combination of these two factors.

The other graphical tool employed in this work is cube plots. This approach consists of plotting the output response, according to the settings which had led to the presented response, in a squared formation. The goal of cube plots is to simply visualize the contribution of each level and parameter in the output response, and to facilitate the manual pattern recognition. A cube plot regarding the experiment example designed in the table 2 is presented in the Figure 7.

Figure 7 – Cube plot of the output response from the experiment designed in the Table 2. The cube approach embraced with a scaled colour map facilitates the visualization and recognition of patterns in the output response according to the factors.



Source: Author.

One may note that the cube plot approach relies on the dimensional characteristic of the total amount of factors in order to be implemented. This means that, designs with more

than four factors can not be easily interpreted with cube plots. A design with two factors can be represented with a square plot and a design with four factors can be represented with several cubes to convey the effect of the fourth factor's levels. For example, if the experiment of the table 2 were structured differently, supposing a 4^2 full factorial design, would have been necessary two plot cubes to represent its features. The first cube would be a representation of the combination of each of the factors' settings at a low (or high) level of the fourth factor, and the second plot would be plotted similarly but under the scope of the remnant factor's level.

Some considerations must be noted in the cube plot analysis. Its original model (ANTONY, 2014) does not contemplate a color map distribution. This was an author's idea to facilitate the manual analysis and pattern recognition of the output response. Furthermore, the dependence of the patterns to be recognized through manual reasoning may lead to misunderstood concerning the true relationships among the factors. Both the interactions plot and cube plots present themselves useful tools to achieve a good notion about the behaviour of an experiment at the frontiers of its volume of settings and to recognize some bias among the parameters interactions. In the cube plot example of the Figure 7, one may just achieve an intuition about how the output response behaves in the extreme of its factor's levels +1 and -1. The inner volume of parameters settings of the cube remains unknown. One could assume a linear relationship among the factors, and basing on this linear model predict all the untried configurations. This approach would work quite well if it were attested that there is no interactions among all the factors, however hardly exists some process or phenomenon whose factors are totally independent from themselves.

Concluded the four phases of design of experiment, one may finally contemplate a reliable intuition of the process or experiment analysed, and profitable settings may be chose. In this work, DOE was employed, at a first approach, in order to map the volume of settings of a sensor, aiming to find the interactions of the device's factors. In a second approach, a full factorial design was expanded aiming to cover a reasonable amount of output responses and the factors which had led to them, so that a dataset with maximum significance could be provided to machine learn models. The cube plots tool was essential to understand and remark the patterns whose ML models could learn and predict.

2.4 MACHINE LEARNING

One may define Machine Learning as a set of computer aided methods that can automatically detect patterns in an amount of data, and deploy the uncovered patterns to take decisions or to predict future states of the data (MURPHY, 2012). Machine Learning approaches relies on the understanding of some phenomenon's features through mathematical models or logical algorithms, so that the coefficients of an algebraic formulation or the weights that penalise decisions can be settled, and thereby the deployment of the trained model leveraged.

Machine learning may be grouped into three main categories: Predictive (or supervised),

descriptive (or unsupervised) and reinforcement learning algorithms. At the first, is expected that, given a vector of independent variable $\mathbf{x}$, the model provides an output $\hat{y}$. In general, $\mathbf{x}$ could be any complex structured object, such as numbers, sentences, graphs, images, etc. and though in theory $\hat{y}$ could be also any of these structures, in practice, most of the methodologies assume $\hat{y}$ as either a finite categorical variable or a real-valued scalar (MURPHY, 2012). When $\hat{y}$ is categorical, the model is known as a classifier, and is known as a regression when $\hat{y}$ is a real-valued. The second main type, descriptive models, does not provide any output given a vector of independent variable. The goal of this approach is to recognize patterns in the data, and to provide relations among them. One interesting descriptive model, namely cluster analysis, groups the independent variable vectors into clusters according to some specific metric, in order to assess the influence of the independent variable in the output response. The last ML main branch is known as reinforcement learning. This methodology is employed to optimize decisions of a software agent in interaction with some system. It induces the software's learning through a reward or punishment signals. In this situation, the input is the constant interaction and response of the software with the system's environment, and the output is a type of decision.

The conceptualization of machine learning in this work concerns the scope of predictive algorithms, specifically regression models and techniques. However, most of the following concepts are intrinsic to any ML category.

2.4.1 Dataset

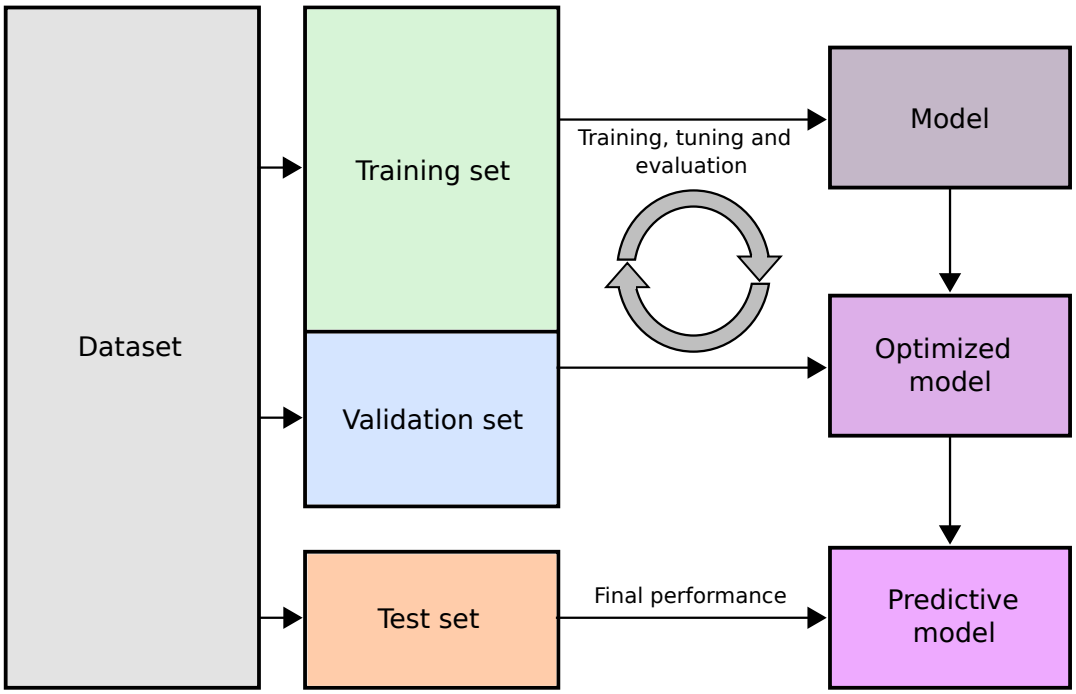
A dataset must be provided to a predictive machine learning model in order to enable it to understand patterns and correlations among the data, and predict the output for unknown observations. A predictive dataset can be stated as:

$$\begin{aligned} D &= \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N} \\ \text{where: } \mathbf{x}_i &= (x_{1i}, x_{2i}, ..., x_{mi}), \\ \mathbf{y}_i &= (y_{1i}, y_{2i}, ..., y_{ji}) \end{aligned} \tag{10}$$

Where $\mathbf{x}_i$ is the i^{th} vector of independent variable of a dataset D whose size is N, and $\mathbf{y}_i$ is the vector of dependent variable respective to the i^{th} vector of independent variable. As mentioned before, both the independent and dependent variable vector can be any complex structured object, such as images, sentences, numbers or even other vectors. The literature also refers to the vectors of independent and dependent variables as feature vector and real output vector, respectively. A dataset of a predictive model is commonly organized as rows and columns. Each column represents the total amount of data of the m^{th} independent variable x_m, with exception of the last columns, which are commonly designated to represent the total amount of data of the j^{th} dependent variable y_j. Each line stands for the i^{th} vector of independent and dependent variable.

An efficient dataset provides a sufficient amount of significant observations in order to train a ML model until its saturation. A wrong assumption at the handling of machine learning models is to assert that the more amount of vectors of independent variable, the better the performance of the model. Large amounts of data may contain noisy samples and outliers, which can impair and ruin the ML model's assumptions. Significant data are samples that better represent the fundamental characteristics of the phenomenon under study. In contrast, poor data can even pretty well represent a local behaviour of a phenomenon with a large amount of data, but the model will fail to remark the global pattern of the phenomenon and hence will fail to perform general predictions. Another factor that may lead to a poor dataset is to have a small amount of observations. In this circumstance, the ML model fails to perform general predictions either due to lack of descriptive information, or overfitting (more details about overfitting can be found in 2.4.6). The lack of information in the dataset so leads to a generic ML model, that its predictions are inaccurate. On the other hand, if the model is overfitted, its predictions will be too bounded with the dataset, so that generalized results can not be achieved.

Figure 8 – Splitting of the dataset during a machine learning application. The original dataset, presented at the left of the figure, is sectioned into two arrangements, one that is deployed to cyclically train, tuning and evaluate the model, and another, that has not been deployed at the learning phase and judges the final performance of the model. The former is named Train set and Validation set, and the last, Test set.



Source: Author.

The process to train and evaluate a machine learning model embraces the management

and splitting of the dataset, as portrayed in the Figure 8. In order to train a model, it is necessary to hold a portion of the dataset out of the training and optimizing process (GOODFELLOW et al., 2016). This portion is referenced as test set, whereas the section deployed to train and optimize is split into training and validation set. The ML algorithm fits its learning design based on the train set observations. The result of this operation is a functional predictive model. However, in pursuance of the best fitting of the algorithm's design into the training observations, enhancements and improvements must be done regarding the handling of the model's singularities and in the model's hyperparameters (more information of hyperparameters in 2.4.8). These enhancements and tuning are carried out founded on the evaluation of the employment of the trained model to predict the output of untrained observations kept by the validation set. Further, the model is again trained with the train set, had its hyperparameters been improved, and the quality of predictions is evaluated once more through the employment of this optimized model to predict the output of the validation set. This cycling process is repeated several times, until the established criteria of predictions' performance are achieved. When this enhancement process is believed optimum, the optimized model must be employed to predict unknown observations, in order to be finally assessed. These unknown observations are kept by the test set, and its functionality is to simulate a real use case for the trained ML model. The assessment of the test set predictions asserts the quality of the model's performance. If the predictions of the test set are not satisfactory, one may step back to the training, tuning and evaluation cycle phase, and attempt to improve the response of the validation set. It is fundamental that the enhancements of the hyperparameters are settled based on the response at the validation set. If the response of the test set is used as a comparison for improvements and tuning, the final predictive model would be fitted to better respond to the test set, and hence the uniqueness of having a set of unknown observations no longer will exist, and the final assessment will not be reliable.

The definition of the contents and structure of the dataset are decisive elements to achieve satisfactory predictions of generic observations. The data must contain the most meaningful information as possible, in order to saturate the training of the algorithm. Besides, a careful planning of the training, tuning and evaluation process must be so carried out, that an optimum final model can achieve a good and reliable performance. When planing the constitution of the dataset, one must consider the splitting dimensions, so that each split has enough data to fulfil its functionality. The standard structure adopted is to reserve 60% of the total dataset amount to the training set, 20% to the validation set and 20% to the test set. In this work, the dataset structure and the enhancements of its data significance were widely exploited in the advancement of the DOE/ML approach.

2.4.2 Regression

Regression is a branch of predictive models that is characterized by the fitting of real-valued independent variables into a learning algorithm, in order to predict a real-valued output, namely dependent variable. Goodfellow et al. (2016) generalize a regression model as a function

f according to the notation 11, where m is the number of elements that consists the i^{th} vector of independent variable $\mathbf{x}_i$ from a dataset represented by the equation 10, and $\mathbb{R}$ is the set of the real numbers.

$$f : \mathbb{R}^m \rightarrow \mathbb{R} \tag{11}$$

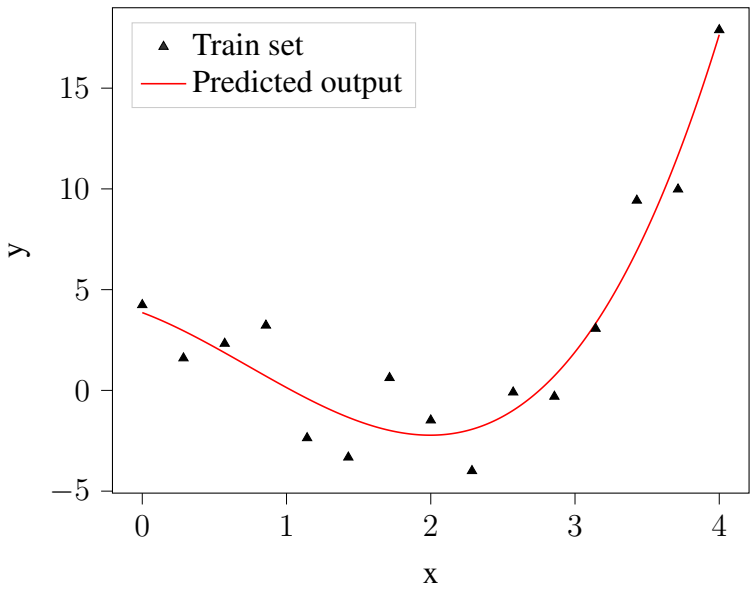
A classical representation of a regression model and application is conveyed in the Figure 9. The Figure represents the fitting of a polynomial regression model into a dataset consisted of just two columns and 15 lines. At the first column is found the unique independent variable that consists the **x** vector, so that m = 1, and the values of x_1 are assigned in the figure as the x-axis values of the triangular black points. At the second column is found the real-valued dependent variable values, which are represented in the picture as the y-axis values of the scattered points. Just a train set was employed in the building and training process of this model, since the nature of the problem is quite simple and its aim is just for exemplification purpose. After the model's hyperparameters have been set, the independent and dependent variable values were deployed to guide the learning algorithm to find the polynomial equation's coefficients that better describe the dispersion of the points. Once the model's metrics are satisfied or no better configurations can be achieved, the fitting of the model is finished and it is ready to be deployed to perform predictions. The targets to be predicted are 300 vectors of independent variable with m = 1 containing real-valued values uniformly ranged from 0 to 4 at the x-axis. The predicted values are assigned to the y-axis through the analysing of the red line.

One may attest that in the Figure 9 the polynomial model was able to construct a curve that generalizes a solution for the given vector of independent variable. Some predictions were pretty akin to the dependent variable's values of the train set, whereas others were quite sparse. A good fitting of model is able to provide generalized and yet accurate predictions to any further dataset that represents the phenomenon under analyses. It is possible to enhance the hyperparameters of the deployed polynomial model so that the predicted values would match perfectly with the train set's real output. But in this circumstance the built model would be too specific trained to the train set, so that it would not be able to perform good predictions of data outer the train set. When this fact occurs, is said that the model is overfitted, and this phenomenon is approached in the Subsection 2.4.6.

2.4.3 Feature scaling

Before the fitting of a machine learning model, a preprocess must be done in the dataset, in order to adapt its elements to be compatible with the algorithm semantic and concept. At the formulation of a regression algorithm model, must be considered whether the learning methodology is affected if the training data contains elements in different numerical scales. Since several machine learning algorithms are parameterized through constants that are simultaneously shared in the fitting of all the independent variables, their different scales might unleash divergent

Figure 9 – Regression model example. Here, a train set consisted of a single independent variable, so that $m = 1$, is fitted into a polynomial regression model. The trained model was deployed to predict observations whose values range from 0 to 4 in small resolution, presented at the x-axis. The prediction result associates the values in the x-axis to the y-axis, through a polynomial curve which best represents the train set distribution for the given hyperparameters. The polynomial regression model employed in this example is provided by the open-source API from Pedregosa et al. (2011).



Source: Author.

learning conditions from the originally formulated, as enlightened by Skiena (2017). The author also states that such divergence leads to unreadable coefficients by the algorithm, numerical imprecision and inappropriate formulations.

In order to empower scaling sensible algorithms to become capable of fitting any nature of real-valued data, there is a need, during the preprocessing phase, to scale the entire dataset. There are two main techniques nowadays employed by the data science community to scale the dataset's features: Standardization (also known as Z-Score) and normalization. The standardization technique operates the scaling of each independent and dependent variable individually, so that, after standardized, the variable's samples have a mean value of zero and a standard deviation of value one. Let $\overline{x}_m$ be the mean value of the m^{th} independent variable's values, and σ_m its standard deviation, the standardization (or Z-Score) technique applied for a value x_{mi}, located at the i^{th} line of a dataset, can be equated as in eq. 12, where $Z(x_{mi})$ is the standardization function.

$$Z(x_{mi}) = \frac{x_{mi} - \overline{x}_m}{\sigma_m} \tag{12}$$

The normalization technique so operates the scaling of a variable, that its values range from 0 to 1. Let x_m^{max} and x_m^{min} be the maximum and minimum value of the m^{th} independent

variable, the normalization function $N(x_{mi})$ for the scaling of a value x_{mi} can be formulated as expressed in eq. 13.

$$N(x_{mi}) = \frac{x_{mi} - x_m^{min}}{x_m^{max} - x_m^{min}} \tag{13}$$

Some of the machine learning models specify in its application programming interface (API) which nature of feature scaling is required to the fitting of the algorithm. The data science community usually employs the normalization scaling when the data of each feature vector element are normally distributed, and standardization when these data are not normally distributed. Most of the regression models that require feature scaling employ standardization.

2.4.4 Evaluation metric

In order to assess the fit of a machine learning model, one must consider which goals the employment of such an algorithm aims, and how one can evaluate its predictions. Concerning predictive algorithms, one might be interested in deploying a reliable metric to compare the similarities between the predicted real-valued output and the real real-valued output (dependent variable). The state of the art for multivariable regression metrics covers four regularly employed techniques, namely Coefficient of determination (also known as R-square, R^2), Mean absolute error (MAE), Mean squared error (MSE) and Root mean square error (RMSE). The evaluations presented in this work were achieved through the coefficient of determination, and thus, the study of the remnant metrics are beyond the present scope. More details about MAE, MSE and RMSE can be found in the work from Härdle e Simar (2015) and Chai e Draxler (2014).

The coefficient of determination is a metric employed to statistically assess the goodness-of-fit of a ML regression model, by the means of contrasting the variance of the predicted output values to the variance of the dependent variable values, as analysed by Allen (1997). The author assumes at first the development of a reasoning to explain the composition of the dependent variable's variance as the sum of the predicted values' variance and the variance of the errors of prediction, as respectively equated in eq.14.

$$\sum_{i=1}^{N}(y_i - \overline{y})^2 = \sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{14}$$

At this expression, N is the dataset size, y is a value of the dependent variable, $\overline{y}$ the mean value of the dependent variable values and $\hat{y}$ is the predicted output. The variance of the dependent variable's value numerically evaluates the spreading of the data with reference to its mean value. The variance of the predicted output, regarding the mean value of the real output, dictates the consistency of the predictions' performance. The variance of the errors of predictions numerically assesses the distance between the true and predicted values. One may remark in eq.14 that, the smaller the value of the variance of the errors of prediction, the more similar is the scattering of the predicted values to the scattering of the dependent variable values, regarding

as reference the mean value of the real output. Given this formulation, Allen (1997) defines as *coefficient of determination* the ratio of the variance of the predicted values to the variance of the dependent variable values as presented in eq.15.

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{15}$$

The equation 15 is commonly expressed in terms of the ratio of the errors prediciton's variance to the dependent variable's variance, through the substitution of eq.14 in eq.15, resulting in eq.16 as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{16}$$

The employment of eq.16 generates an interpretable real-valued number in the range from 0 to 1, so that $0 \leq R^2 \leq 1$. The closer the value of R^2 is to 0, the less the predicted output's variance is associated with the dependent variable's variance, and according to eq.14, this variance is thus due to the error of prediction's variance between $\hat{y}$ and y. On the other hand, the closer the value of R^2 is to 1, the greater the similarity of the predicted output's variance is to the dependent variable's variance, and hence less is the error of prediction's variance. One may note in eq.16 that R^2 can reach negative values. This circumstance implies that the model's performance is worst than just to assume each value of $\hat{y}$ as being the average value $\overline{y}$ of the real output values.

The coefficient of determination is a practical and generic evaluation metric that can be deployed to assess the performance of any regression model, so that it enables the comparison of different regression algorithms concerning how well each of them fits the data (ALLEN, 1997). This metric was widely deployed in this work, in order to assess the performance of the employed algorithms and how its accuracy varies according to the change of settings of the algorithms' hyperparameters.

2.4.5 Bias and variance

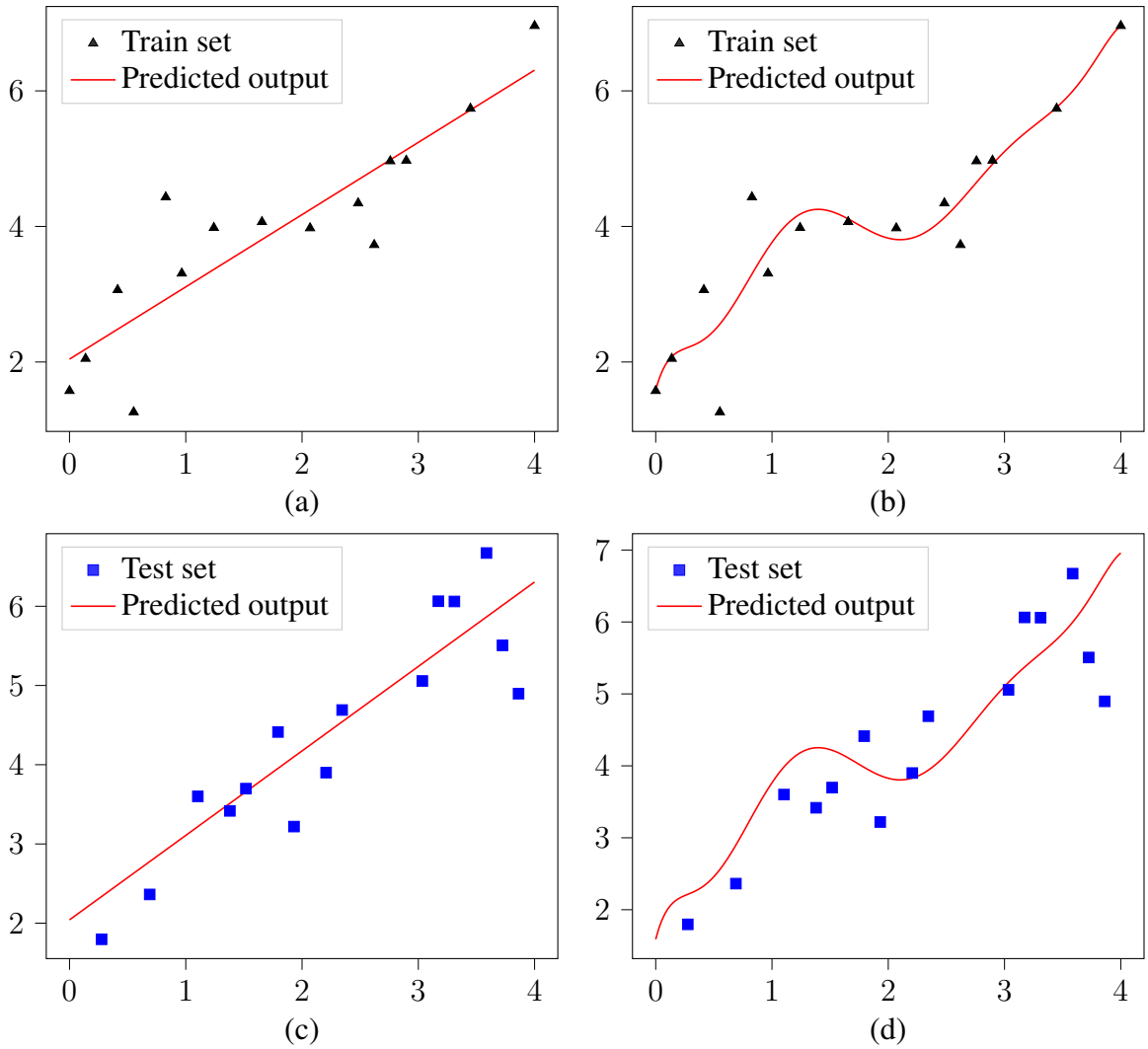
Concerning the fitting and performance of a machine learning model, there are two main concepts that influence directly the model's accuracy and reliability, namely overfitting and underfitting, which are explained through two statistical notions: Bias and variance (GOODFELLOW et al., 2016). These four approaches are intrinsically related. To find the sweet spot among them assures satisfactory predictions. Skiena (2017) defines the bias error of a ML model as the incorrect assumptions taken by the algorithm about the relationship of the train data that builds the model. The author also states variance error as the performance sensitivity

of the built model regarding datasets different from the train set. The Figure 10 exemplifies these concepts. All the models deployed in the figure are provided by the open-source API from Pedregosa et al. (2011)

The example is carried out supposing a dataset consisted of just one independent variable, so that m = 1, and one dependent variable. The dataset is split just into train and test set. A linear and polynomial regression are trained and deployed to predict the respective output for one independent variable, which consists of observations that numerically range from 0 to 4 in small resolution. At Figure 10.(a) and (c), the linear regression is trained with the train set, and the algorithm fits a straight line through it, so that the line strives to represent the relationship among the data. The Figure 10.(b) and (d) employ the same train set used in the linear regression, in order to train a polynomial regression model, which strives to fit a polynomial curve into the train set, aiming to represent the patterns among the data. Observing the Fig.10.(a), one may remark that the straight line does not have the flexibility to accurately represent the data's relationship, because the model predefines the assumption that the train set values are linear correlated among themselves as a straight line. This predefinition of how the ML model is going to fit the training data is known as model bias, according to the definition made by Skiena (2017). Since the linear regression assumptions are hardily softened, it is said that the model has a high bias. At Fig.10.(b) the polynomial regression is built with a high polynomial degree level, so that the algorithm almost interpolates the train set values. And since the model (when built with a high polynomial degree level) roughly does not state assumptions about the train set fitting, it is said that the model has low bias. The coefficient of determination's value of the built linear model from Fig.10.(a), for the predictions of the train set's dependent variable is 0.78, whereas the same metric applied for the same predictions performed by the polynomial model from Fig.10.(b) achieved the value of 0.86.

The figures 10.(c) and (d) presents the performance of the same built models from Fig.10.(a) and (b) respectively, predicting the output for the same independent variable consisted of numerical observations ranging from 0 to 4 in small resolution, but now the untrained observations from the test set are the reference to assess the deterministic model's accuracy. At the Fig.10.(c), one may remark that, though the presented fitting is not the best representation, yet the straight line of the linear regression is capable of describing the test set values as well good as the train set values. Actually, its coefficient of determination for the predictions of the test set real output is even higher than the value achieved for the training set, reaching 0.81. The assessment of the linear regression model leads to the acknowledgement that the performance of the algorithm is not significantly sensitive to fluctuations of dataset, since the coefficient of determination value achieved with the predictions of the training and test set are roughly the same. The sensitivity of a built model's performance to different datasets is known, according to the definition stated by Skiena (2017), as variance, and thereby is said that the linear regression model has low variance. On the other hand, at the Fig.10.(d), one may remark that since the fitting of the model's polynomial curve into the training samples was too conservative to the training

Figure 10 – Linear and polynomial regression built models to predict the real output for a dataset consisted of one independent variable. All the four red splines represent predicted outputs for one independent variable consisted of observations numerically ranged from 0 to 4 in small resolution. The Fig.(a) presents the fitting of the linear regression straight line into the train set. The strong regularization of the model's assumptions of the data characterizes it as owning high bias. At Fig.(b) is shown a high level degree polynomial regression built model, so that the fitting curve almost interpolates the train set values. Since the model's assumptions of the data's relationship are pretty flexible, it is said that the algorithm has a low bias. At Fig.(c) the very same built model from Fig.(a) is presented, but the assessment reference is the test set. Since the accuracy of the predictions of the test set are visually quite similar to the predictions made for the test set, is said that the linear regression model has low variance. At Fig.(d), the same built model from Fig.(b) is shown, but as in Fig.(c), the assessment reference is the test set. Since the regression's polynomial curve is too fitted into the train set, it visually achieves worse accuracy of predictions at the test set, and hence it is said that the model has high variance.



Source: Author.

set characteristics, the employment of such a curve to represent the patterns of an unknown set of data, in the case of Fig.10.(d) the test set, does not present the same good predictions as it did for the train set. Indeed, the coefficient of determination of the prediction of the test set real output made by the model from Fig.10.(d) reached the value of 0.71. This assessment achieved by the means of the coefficient of determination metric attests numerically the visual belief that the performance of the predictions of the test set was worse than it did in the train set. Hence, is said that the model from Fig.10.(d) has a high variance.

The bias and variance of a built model are defined by the intrinsic construction of the fitting algorithm and by the choice of the hyperparameters (see 2.4.8). Neither the linear nor the polynomial regression built in the example of Fig.10 has an optimum setting regarding the caution to handle with bias and variance errors. In fact, a trade-off between these two concepts must be so considered in the building of a model, that the learning algorithm does not be too regularized with data assumptions, leading to a model with high bias, but neither comply with too soft assumptions, so that the fitting is entirely dictated by the training data, leading to a model with high variance. The ideal model fitting would have low bias and low variance, in such a manner that the algorithm would be capable of learning a quite flexible fitting with the train set, and this fitting would be generic enough to depict any further batch of data from the phenomenon under study. However, since the learning method is immutable, one can manage the trade-off between bias and variance by means of tuning the hyperparameters, in order to find a sweet spot between the model's bias and variance. Hence, the built ML algorithm would be empowered to perform accurate predictions for different sets of data and reliably deployed in future applications.

2.4.6 Overfitting, underfitting and generalization

As mentioned in Section 2.4.5, the well-fitting of a model relies on two main ML concepts, namely overfitting and underfitting. Goodfellow et al. (2016) describe these phenomenons as the two central challenges in machine learning. The authors state underfitting as the inability of a model to reach a low error value in the prediction of the training set, and asserts overfitting as the inability of a model to achieve similar error performance between the prediction of the training set and the prediction of the test set, i.e. the error of prediction reaches significance different scales for each set of data predicted. The authors also convey that these two concepts are explained through the statistical notions of bias and variance, approached in Section 2.4.5. When a model is too conservative in its assumptions of the data relationship, such as the linear regression of the Figure 10.(a) so that it fails to well describe the data distribution due to high bias, the ML algorithm fails to well perform in the train set prediction, and it is said that the model is underfitted. On the other hand, it is said that the model is overfitted when the algorithm is too softy in its assumptions and strives to achieve accurate performance in the train set prediction through the memorization of the train set features, as remarked in Figure 10.(b), so that it fails to describe the test set due to high variance.

Goodfellow et al. (2016) agree that the goal of a machine learning algorithm is to perform well at the predictions of new, and yet not seen, observations. The authors define the ability of a model to achieve good predictions at a yet unlearned dataset, as generalization. A good and reliable ML built model is capable of well learning the relationship of the train set, reach a good performance in the prediction of the training samples, and further be employed to predict unknown observations and achieve similar good predictions in it as well as in the train set, complying with the definition of a generalized model. The achievement of a generalized model can be pursued through the choice of a learning algorithm that is best able to recognize the patterns of the data, tunning the hyperparameters, cross-validation techniques, ensemble methods and other approaches enlightened by Goodfellow et al. (2016).

A particular circumstance that regularly leads to underfitting is the training of a ML algorithm with a small dataset. The handling of this situation requires caution and it is approached in Section 2.4.7.

2.4.7 Small dataset features

The fitting of a machine learning model relies on the association between the learning algorithm and the adjustment of its parameters to mark the patterns of the data. However, when the dataset does not provide a large amount of observations, the degree of freedom of a model (i.e. the number of parameters of a model regarding the number of observations in the training set) increases as a result of the settlement of the model's unsoftened assumptions, i.e. large bias, leading to underfitting and consequently decreasing the accuracy of the predictions (ZHANG; LING, 2018). In fact, it is a well consolidated foundation the notion that a machine learning algorithms' performance improves with the increase of the amount of meaningful data of the training set, as reported by Schmidt et al. (2017) and Faber et al. (2016).

Nevertheless, the employment of a small dataset is inevitable at certain situations, once that the cost of each observation is expensive and a greater amount of data is not affordable or simply impossible, as presents Zhang e Ling (2018) and Kamath e Fan (2018). Thereby, in order to empower the deployment of ML algorithms in such a small set of observations, one may approach the problem employing different branches of ML strategies to enhance specific methodologies in the management of the problem, and hence rendering the model competitive. For instance, one may improve the data's significance of the training set through sampling the observations by a DOE methodology, and deploy the evaluating technique of Leave-One-Out (LOO) in order to accomplish an optimization phase of the model's building, as did Kamath e Fan (2018). Another strategy to outperform the achievement with small datasets is through the softening of the model's bias, by the means of ML ensemble techniques, as studied by Webb e Zheng (2004).

Since the nature of the circumstance approached in this work handles with small datasets, all the three strategies mentioned above were employed into the resolution of the problem of this work.

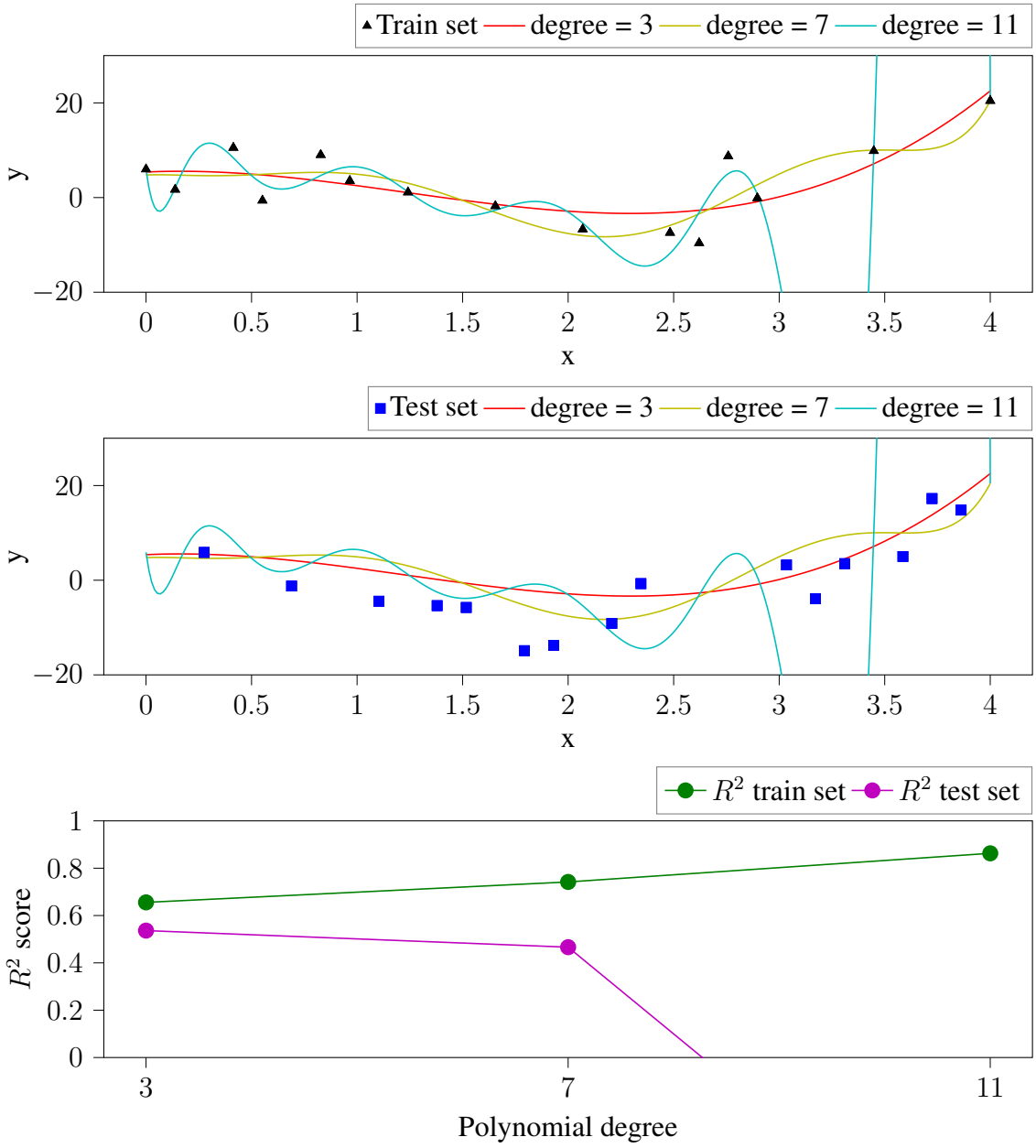
2.4.8 Hyperparameters

The hyperparameters of a machine learning model are the settings of the learning algorithm that one can set, in order to control the learning algorithm's behaviour, as assert Goodfellow et al. (2016). The polynomial regression of the Figure 9 has just one unique hyperparameter, namely the level of the polynomial degree itself. However, most of the ML models have several hyperparameters that enable the customisation of the learning approach. Such customisation is crucial to fit an algorithm into a specific dataset, which has singularities that must be considered in order to empower the interpretation of the data's patterns by the learning methodology.

The setting of the hyperparameters must be designated with caution. In fact, such is the importance of this procedure that a hyperparameter can not be adapted by the learning algorithm itself, i.e., it can not be learned (GOODFELLOW et al., 2016). The reason for this statement is that a learning methodology will always try to fit the training set in such a manner that the best performance might be achieved in its predictions. Consequently, the algorithm would improve the hyperparameters values until the predictions of the training set reach an optimum of similarity. But as it is demonstrated in Figure 11, the careless enhancement of the polynomial degree would lead to overfitting. The Figure 11 demonstrates how the training and testing performance of a polynomial regression varies according to the enhancement of its hyperparameter, namely the polynomial degree. The regression model was trained with a dataset consisted of just one real-valued independent variable with values that range from 0 to 4, represented in x-axis, and a real-valued dependent variable represented in the y-axis. The algorithm was deployed to predict the response for 300 real-valued observations within the range of 0 to 4, distributed in small resolution. At Fig.11 top, one may remark the increase of complexity of the polynomial curve at the description of the train set, given the increase of the polynomial degree. The Fig.11 bottom presents the R^2 value for each different hyperparameter level in the train and test set. One may note that, the more squiggly the predicted curve for the train set, better its performance in the R^2 metric. Nonetheless, since the curve's complexity is fitted for the train set, the higher the polynomial degree, the worse the performance in the test set, as one may visually observe in Fig.11 middle and numerically attest in Fig.11 bottom. In fact, the careless tuning of the hyperparameters can lead to discrepant situations, where the predictions achieve excellent results for the train set, and catastrophic for the test set, achieving even negative values for the R^2 metric, as occurs in Fig.11 bottom when the polynomial degree is set to 11.

The process to determine the values of the hyperparameters is carried out during the cyclic phase of the model's tuning and evaluation, assessing its performance predicting the validation set, as detailed in 2.4.1. At this phase, one may remark and assess, through chosen metrics, how the model's efficiency of the prediction of the validation set varies according to the changing of the hyperparameters. Goodfellow et al. (2016) depict this phase as the using of the validation set to "train" the hyperparameters. Once that the algorithm's performance at the

Figure 11 – Polynomial regression built model employed to predict the real output of a dataset consisted of one independent variable. All the coloured splines from the Fig. top and middle represent predicted outputs for one independent variable that consists of observations numerically ranged from 0 to 4 in small resolution. At fig. top, it is presented the built model predicting the very train set given different hyperparameter's levels. In Fig. middle, is found the response for the same built model from fig. top, but here deployed to predict the response for the test set. At Fig. bottom is presented the R^2 value for each different hyperparameter level in the train and test set. The polynomial regression model employed in this example is provided by the open-source API from Pedregosa et al. (2011).



Source: Author.

prediction of the train and validation set is satisfactory, the generalization of the model may be assessed by means of the deployment of the model to predict the test set.

2.4.9 Regression models

There are several different logics and principles that guide the learning algorithm of a machine learning regression model. Each technique has its own intrinsic methodology and particular aims during the fitting process, in order to better describe the data's relationship and hence perform good predictions. In this section, brief explanations of the principles and features of four regression models are found. These models were employed in the experiments of this work.

2.4.9.1 XGBoost

XGBoost is a tree boosting based algorithm, developed by Chen e Guestrin (2016). A tree-based ML model is consisted by the partitioning of the input space, followed by the assignment of a simple descriptive model (for instance a simple constant) to each split region (BISHOP, 2006). Tree-based algorithms can be comprehended as a collection of descriptive models, in which only one description is assigned to perform predictions for a given vector of independent variable in the input space. The assimilation process of each vector of independent variable to a descriptive model is carried out by means of a sequential decision process through the traversal of a binary tree, so that it is characterised by the splitting of two branches at each node. The boosting strategy consists of building a predictive model by the ensemble of "weak learner" algorithms (MURPHY, 2012). Each weak learner is trained in sequence employing weighted versions of the dataset, where the weighting of the data is associated according to the performance of the previous learner. In particular, the worse the performance at earlier rounds, the greater the weight value for the next learner. Thereby, a tree boosting algorithm consists of an ensemble of tree-based models.

The XGBoost model performs the boosting through the Gradient Boosting methodology (CHEN; GUESTRIN, 2016), so that it generalizes the boosting algorithm to handle with generic loss functions, and hence regression (MURPHY, 2012). XGBoost also innovates among general tree boosting algorithms by computing a penalty term in the minimisation of the loss function, in order to reduce the model's complexity, and thereby preventing overfitting (LUCKNER; TOPOLSKI; MAZUREK, 2017). In addition, the model provides a scaling parameter in order to control the impact of a single tree in the final score, and implements a column subsampling technique to avoid overfitting and improve the computation speed (CHEN; GUESTRIN, 2016).

2.4.9.2 Gaussian process

Gaussian process (GP) regression is a probabilistic based model. Bousquet, Luxburg e Ratsch (2004) define GP as a collection of random variables, in which any finite batch has a consistent joint Gaussian distribution. GP can also be considered a descending extension of multivariate Gaussian distributions to infinite dimensions, as states Kamath e Fan (2018). The authors justify the employment of Gaussian process in machine learning, if it is assumed a

dataset of size N as a single point from a N-variate Gaussian distribution. GP is totally described by a probabilistic model underlain in a mean function $\psi(\mathbf{x})$ and covariance function $\kappa(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathbf{x}$ is a vector of independent variable, so that the model is a generalization of Gaussian distribution in order to process functions (BOUSQUET; LUXBURG; RATSCH, 2004). Given the input data, the algorithm infers a distribution over functions, and rely on these functions to perform predictions (MURPHY, 2012).

The covariance function of the GP model structure is elementary in the training of the learning algorithm, since it codifies the assumptions of the function that will be learned, by means of the determination of the definition of the degree of similarity among vectors of independent variables, so that nearby vectors of independent variables have high correlated outputs (SU et al., 2014). In Gaussian process the mean and covariance functions are hyperparameters, so that prior assumptions and beliefs about the structure of the model may be specified. The covariance functions also are regarded as *kernels*, and it is common that several definitions of them can be found in conjunction with the GP's API.

2.4.9.3 K-nearest neighbor

The K-Nearest Neighbor (KNN) model is a nonparametric based algorithm, i.e., it does not take any prior assumption of a learning model function before any data be observed, but it designs its complexity as a function of the training data (GOODFELLOW et al., 2016). The training of the learning algorithm consists of simply storing the training set and the hyperparameters' values. For regression, given an input independent variable vector, the model searches in the training set for the K-nearest vector of independent variables and returns the mean value of their respective dependent variables' values, as explained by Kramer (2013). The author clarifies the idea justifying that since the local neighbors of a vector of independent variable $\mathbf{x}_i$ are expected to have similar continuous dependent variable's value y_i, the prediction for an unknown variable $\mathbf{x}_j$ must has similar value to the value of its neighbors, and such similarity may be modeled through averaging. The neighbor's closeness is calculated through a distance metric defined as a hyperparameter, since each dataset may have different features that can be better related by a personalized distance metric. It is common that the KNN's API provides several options of distance metrics.

The KNN methodology induces local constants responses. In order to smooth the prediction function, Kramer (2013) presents the weighting of the neighbor's similarity as an alternative to avoid plateaus prediction regions. The idea is that closer independent variables shall contribute more for the final prediction value, leveraging different possible optimizations concerning the data relationship.

2.4.9.4 Support vector machine

Support vector machine (SVM) is a linear based ML model. The algorithm computes a linear regression through a high-dimensional (or infinite-dimensional) mapped feature input space (WAUTERS; VANHOUCKE, 2014). The ML model exploits the fact that only the inner product is used in the calculations, to employ kernel functions, i.e. different definitions of the inner product, in order to enable nonlinear regression performances (BALABIN; LOMAKINA, 2011). Since the relation between the independent and dependent variables are hardily linear, the model defines a new Euclidian space as a result of a personalized inner product (also regarded as mapping function) (LUU et al., 2009), aiming to find a hyperplane that better describes the independent and dependent variables' relationship in the new dimension space, and based on this description it performs predictions (WAUTERS; VANHOUCKE, 2014).

SVM may be deployed to solve classifications and regression problems. For regression, the learning algorithm is built through the penalization of the prediction's errors, in order to minimize a cost function (BALABIN; LOMAKINA, 2011). The fitting of the linear regression in the high-dimensional feature space is controlled by the setting of two mainly hyperparameters, which regularize the performance at the cost function, so that the model's overfitting aspects can be managed, and the best fit reachable.

2.4.10 Ensemble methods

Ensemble methods are a collection of algorithms that try to construct a set of learning models and combine them (ZHOU, 2012), aiming an enhanced predictive performance. An ensemble is consisted of a certain amount of base learners that are base learning algorithms trained with training sets from the same dataset. Zhou (2012) distinguishes ensemble methods regarding their composition of base learning algorithms, stating as homogeneous ensembles, methodologies that employ just one type of base learning algorithm in its performance, and heterogeneous ensembles, as techniques that employ more than one type of base learning algorithm. The author also notes that the generalization of ensemble models, i.e. the capability of a model to achieve good predictions of a yet unlearned dataset, is often much higher than those achieved by single base learners, since the technique is designed to boost the performance of the base learners. It is a general belief that, in order to achieve a good ensemble, the choice of the base learning algorithms must be carried out with caution, so that each base learner should be as accurate as possible, and as diverse as possible (ZHOU, 2012).

There are several ensemble learning algorithms and methodologies, such as bagging, boosting, combination methods and variants of these techniques, depicted by Webb e Zheng (2004). Each of them performs different strategies regarding the training and predictions phase, in order to combine the features of each base learner. Since that in this work it is employed just the *combination methodology* of ensemble learning, the other strategies are beyond this scope and will not be approached. More details about them may be found in Zhou (2012) and Webb e

Zheng (2004).

Combination learning usually relies on heterogeneous ensembles, so that rather to try to find the best single learning algorithm, combination learning states the final prediction basing on the predictions of the base learners. The most popular and fundamental combination method for regression is known as *Averaging*, whereas *Voting* stands for classification (ZHOU, 2012). To comprehend Averaging, it is supposed a conjunct T composed of individual base learners $\{h_1, h_2, ..., h_T\}$, so that the output of h_i, given a vector of independent variables $\mathbf{x}$, is the predicted value $h_i(\mathbf{x}) \in \mathbb{R}$. The final prediction $H(\mathbf{x})$, for a single vector of independent variables $\mathbf{x}$, in combination with all the base learners, is the simple average of each individual learner, such as describes the equation 17.

$$H(\mathbf{x}) = \frac{1}{T}\sum_{i=1}^{T} h_i(\mathbf{x}) \tag{17}$$

It can be proved that the error of prediction of Averaging ensemble is smaller by a factor of T than the averaged error of the individual learners (ZHOU, 2012). However, this proof is only true if there exists no correlation between the trained learners, which is not the case for the ensemble methodology, where the learners are being trained in the same dataset to solve the same problem. Thus, the error reduction by a factor of T is hardly reached. Nonetheless, the simplicity and effectiveness of Averaging is widely acknowledged by the data science community. Indeed, the main advantages to employ combination methodologies in machine learning applications is that they are more accurate, has lower bias and lower variance than single base learners (ZHOU, 2012).

3 MATERIALS AND METHODS

To achieve the characterisation of the device under study through the DOE and ML approaches, a conceptualization of the combination these two methodologies, in order to complement each other, was pondered. Experiments were performed based on the DOE theory, and fundamental conclusions regarding the inclinometer's behaviour were so achieved that a significant data set could be extracted from these experiments, in order to train a machine learn model to fully characterise the device's features for a delimited scope of parameters' combinations. In order to perform the experiments, a reliable experimentation methodology and communication infrastructure were developed for the presented use case.

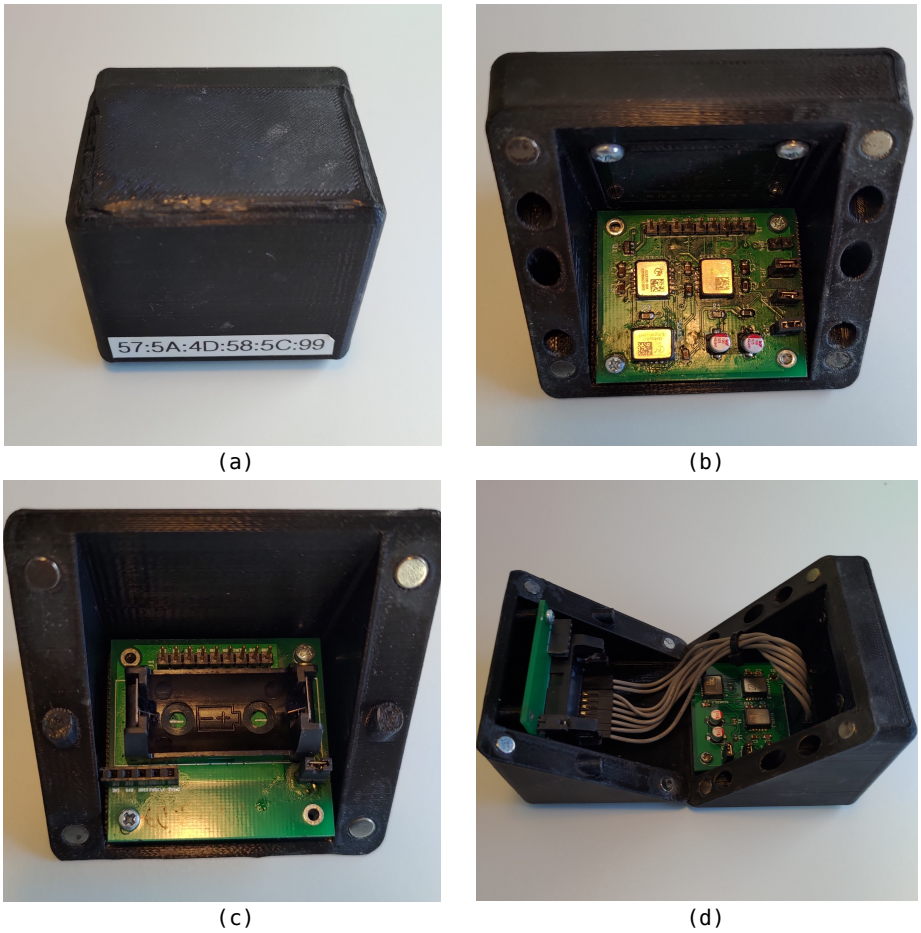
In this chapter is described the equipment and methodology employed to carry out the experimentation and analyses. First, is presented an introduction of the device's functionality, its parameters and the variety of its noisy response accordingly with the parameters' settings. Hereupon is explained the communication infrastructure developed to dynamically perform experiments and assess the data acquired by the sensors. Further, an evaluation of some singularities of the inclinometer are presented. In sequence the experimentation methodology founded on DOE theory and data treatment is detailed. Afterwards, the inclinometer's design of experiments is presented, and following, the results of this evaluation's model are presented, in order to ponder conclusions on which the machine learning relies to be accomplished. To conclude, the ML approach is conceptualized into the inclinometer's DOE.

3.1 THE INCLINOMETER DEVICE

The inclinometer smart sensor device deployed in this work is a project developed by the team of Large-Scale Metrology under supervising of the Chair of Production Metrology and Quality Management, WZL at RWTH Aachen University. In order to map the structural thermal displacement of a machine tool, the Large-Scale Metrology team has designed an inclinometer smart sensor able to be robust, wireless, dynamically adaptive to different situations, configurable, and accurate, complying with the standards of smart sensors (Section 2.2). To carry out this accomplishment, the team had relied on the assembly of already existent electronic modules to perform the measurement itself and the management and broadcast of the data. The device developed is presented in Figure 12.

The sensor is projected to read, through the active acceleration felt by a sensing mechanism (Subsection 2.1), the static tilt angle of two orthogonal axes, namely x and y, and to read the dynamical acceleration for three degrees of freedom (DOF). The acceleration of the static tilt is measured by two MEMS ICs SCA830 from Murata (2020c) , whereas the 3-DOF dynamical acceleration is measured by the MEMS IC SCA3300 also provided by Murata (2020b). Both these modules are found in Figure 12 (b). In this work, only the functionalities

Figure 12 – The pictures in the figure portray the sensor device explored in this work. Fig.(a) presents a visualization of the sensor ready to be used. In fig.(b) is found the MEMS IC module. Fig.(c) shows the electronic board that assembles the battery and the BLE microcontroller. Finally, in fig.(d), one may behold the entire assembly of the electronic modules, which are connected through SPI bus line.



Source: Author

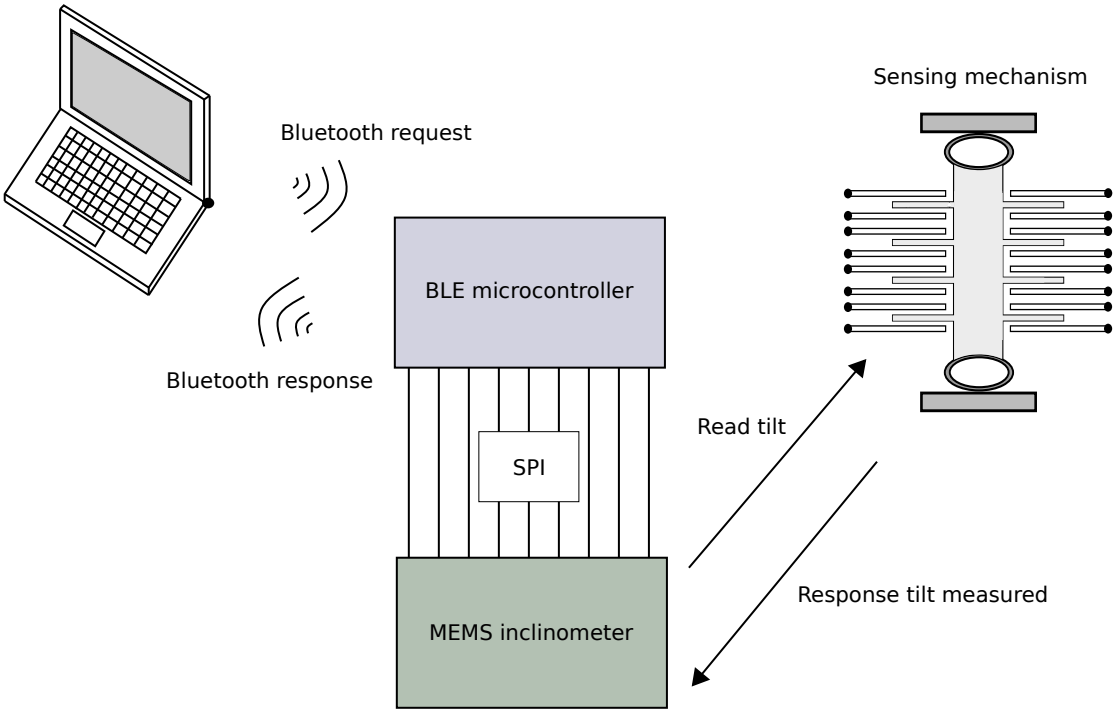
concerning the static tilt modules were explored. The characterisation of the dynamical sensor was not accomplished.

After the tilt measurement is performed, the information is conveyed through wired SPI communication protocol to the BLE microcontroller CYBLE-212006-01 from Cypress (2019), where the data is processed and broadcast through Bluetooth 4.2. The BLE microcontroller is part of the assembly from Figure 12 (c). The assembly complex is energized by a single 3.6V battery, that is attached to the battery structure on the assembly of Fig.12 (c). The electronic modules are kept by a 3D printed case. This housing has magnets whereby the case can be magnetized on the surface of a machine tool.

The BLE microcontroller propagates periodically notifications packages, in order to be

recognized by some central device and establish the Bluetooth connection topology (Subsection 2.2). Once connected, the central device (in this work a personal computer) sends a configuration package, in order to dictate to the controller the parameters and configurations of the measurement that are required. The BLE microcontroller process these parameters, and establishes a routine of measurement, according to the settings received, with the MEMS modules through SPI communication protocol. The sensor ICs then perform the measurement of the tilt position that the sensing mechanism is found regarding the gravity vector (Subsection 2.1). The measurement data is preprocessed and sent back to the BLE microcontroller through SPI. The controller, in turn, processes the data and broadcast them to the central device through the Bluetooth connection. The Figure 13 summarizes this schematic of communication and data transferring.

Figure 13 – The figure presents a schematic of the communication among the inclinometer device's modules. At first, a measure request is made by the central device, which in the case of this work is a personal computer, to the inclinometer. The request is interpreted by the BLE microcontroller, and this, in turn, dictates to the MEMS inclinometers a routine of measurement's reading. The MEMS read the active acceleration interpreted by the sensing mechanism, and this, in turn, return the measurement through SPI to the BLE microcontroller, which broadcast the data received to the central device that had requested it.



Source: Autor

3.1.1 Inclinometer's parameters

The procedure of establishing a Bluetooth connection with the inclinometer device is carried out through an API written in python by the researchers from the team of Large-Scale Metrology. By the means of this interface, the user can specify some parameters regarding the Bluetooth connection and the routine of measurement. Since the Bluetooth related parameters do not interfere with the quality of the measurement, they are not approached in this work. The settlement of the routine measurement's parameters dictates some details of the procedure performed by the MEMS ICs, and are enlightened as following.

- *Operation mode*: There are three routines of measurement that can be chosen, namely *Standard*, *Repeated* and *Continuous*.
  - *Standard*: When the *Standard* is settled, the BLE microcontroller specifies a period of time in which the MEMS ICs are going to keep reading and collecting samples of the actual tilt placement. The reading is performed with a certain frequency regarding the period of time chosen. For each sample collected, the MEMS ICs are going to read a certain number of times the sensing mechanism, and return as a single sample the average value of the amount consulted. After the period of reading time is ended, the BLE microcontroller averages the value of all the samples collected in this interval, sends the averaged value and the standard deviation of the averaged samples to the central device that requested the reading. The inclinometer device then establishes an idle operation mode, where the sensor does nothing and waits for the next instructions. An important aspect to observe is that the Bluetooth connection remains established while the inclinometer is in idle mode, so that new requests of standard reading can be performed without the need to reconnect the Bluetooth protocol. The Algorithm 2 in the Appendice A summarizes the procedure of the *Standard* operation mode.
  - *Repeated*: The *Repeated* mode follows the exact same routine as the *Standard* mode, but instead of being set to an idle mode after have sent the averaged value to the central device, the same measurement routine is executed again. The MEMS are going to read samples within a period of time, with a certain frequency, where each sample is an average of a certain number of consultations of the sensing mechanism, and after this interval of measuring is ended and the averaged value of the samples collected in this period is sent, along with the standard deviation of the averaged samples in this interval, then the entire procedure is rerun. The repeated procedure is only ceased when the Bluetooth connection is set to be broken. The Algorithm 3 in the Appendice A summarizes the procedure of the *Repeated* operation mode.
  - *Continuous*: The *Continuous* has a different principle of routine. When set to this mode, the BLE microcontroller dictates to the MEMS ICs to collect samples at a certain frequency (this frequency regarding one second), and return the exact sample

read to the central device, continuously. For each sample, the MEMS ICs are going to consult the sensing mechanism a certain amount of turns, average the values observed and return this value as a single sample. This sample is sent to the central device and the MEMS ICs are scheduled to perform another measurement according to the frequency settled. This procedure is repeated continuously until the Bluetooth connection is set to be broken. An interesting point to be remarked is that each sample sent is not an average of a certain amount of samples collected during a certain period, and hence no standard deviation of the samples collected in each measurement procedure is calculated in this mode. The Algorithm 4 in the Appendice A summarizes the procedure of the *Continuous* operation mode.

- *Interval*: Period of time within the MEMS ICs is going to collect samples of the actual tilt placement. This parameter is only used in the *Standard* and *Repeated* operation mode.
- *Frequency*: Rate of sampling tilt measurements. For the *Standard* and *Repeated* operation mode the frequency of reading the MEMS ICs is regarding the interval selected. For the *Continuous* mode the rate of sampling is regarding one second.
- *nValue*: For each sample required to be read by the MEMS ICs, a certain number *nValue* of observations of the sensing mechanism are going to be done by the MEMS and the average value of the samples is going to be returned as a response for the single sample required.

Another parameter was observed empirically during cycles of measurement requirements for the *Standard* mode. It is possible to architect an algorithm to read the inclinometer in the *Standard* mode sequentially. In order to accomplish that, one may use the API interface provided by the team of Large-Scale Metrology to write a loop of standard measurement while the Bluetooth connection stands, and define a time of interval between the standard measurements, so that each routine of standard measurement will be started after a certain amount of time be fulfilled with reference of the last finished measurement. Since the name of the python function which regularizes this difference of time is named as *time.sleep()*, this name was assigned to refer to this period of time. The definition of this parameter is stated as follows.

- *Timesleep*: This is an experimental parameter, and only occurs when performing routines of measurement in the *Standard* mode. The parameter is the amount of time between the requests of measurement routine made by the central device.

The possibility to customize the routine of measurement through the settlement of these several parameters is one of the goals of the inclinometer developed, once that one may adapt the sensor for each desirable situation. By the regularization of the *Interval*, *Frequency* and *nValue*, the power consumption can be manageable according to the necessary sampling rate of the application. The parameter *interval* can be manageable with a step resolution of 1 millisecond, the parameter *frequency* with a step resolution of 1 Hz (for any operation mode), the parameter *nValue* with 1 unity and the parameter *timesleep* can be configured with a step resolution of any real-valued number. The *Operation mode* can be settled to better comply with

the use case. And the *timesleep* can be configured to regularize the rate of execution routine in the *Standard* mode. Since it is countless the number of combinations of parameters, simply analysing the performance of the device for each combination is not a practicable option. A reliable and feasible methodology to map the inclinometer's performance for a delimited volume of settings is the subject of this work.

3.1.2 Inclinometer's response

The inclinometer developed provides as output for the request of tilt reading the active acceleration of the sensing mechanism regarding the gravity vector, as detailed in Subsection (2.1). Since the device is projected with two MEMS inclinometers, placed orthogonality from each other, it is possible to apply the Equation 9 to map, through the acceleration response, the inclination of the sensor device of the two orthogonal axes, namely x and y.

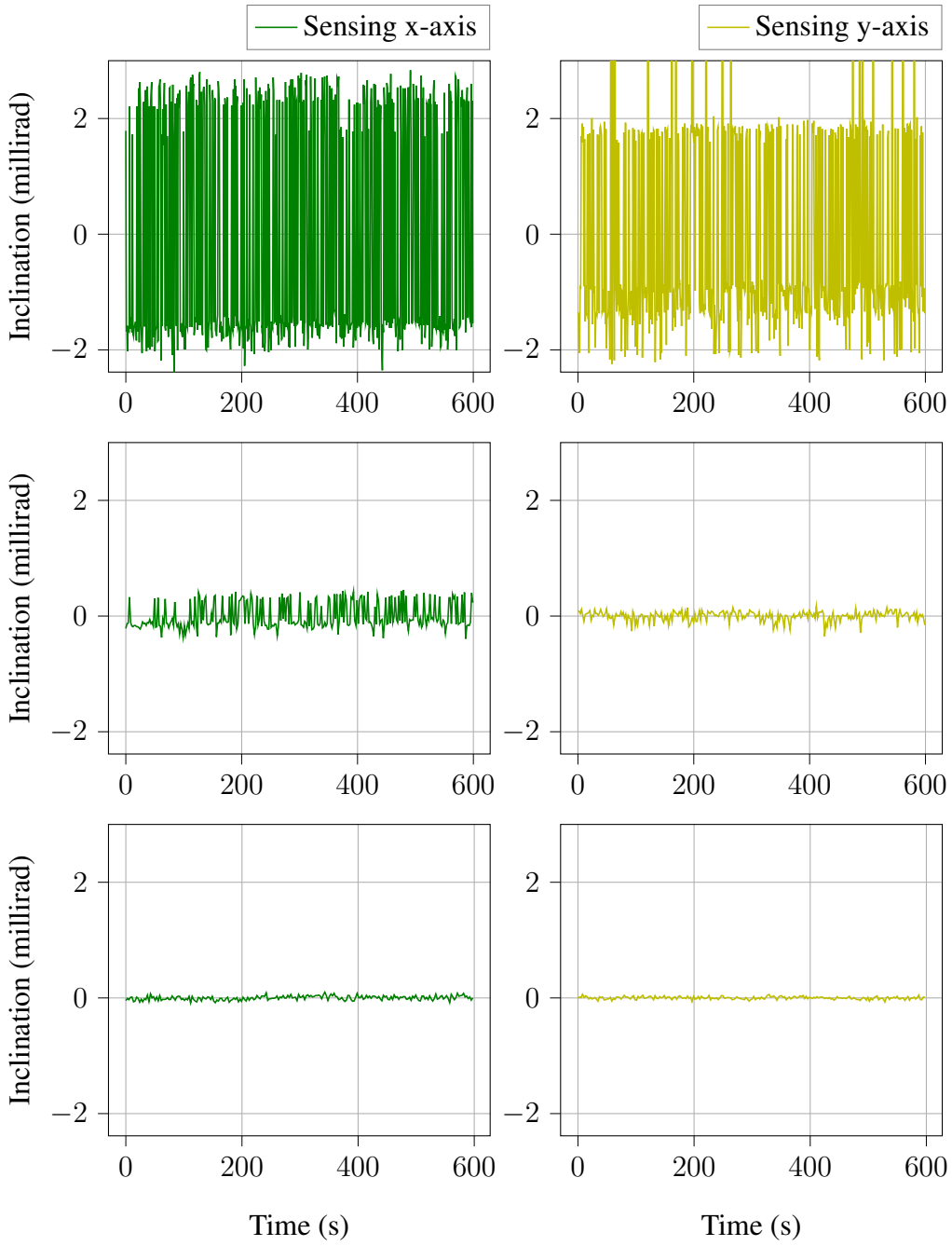
As detailed in Subsection 3.1.1, several configurations concerning the routine of the device's measurement can be settled, and the performance of the tilt measured's response varies according to the combination of these parameters. The Figure 14 presents the sensor's response for each axis of measurement at three different configurations of parameters. Each horizontal couple of graphics exposes the response of each of these three combinations of parameters. For each set of parameters, the sensor was left laid still on a steady and assumed plain surface, reading its inclination for several minutes. Ideally, the device's response for this scenario should relate 0g (0 radians when applying the Equation 9), as depicted in Figure 2, for any set of parameters. However, since the manufacture of the inclinometer device is handmade and MEMS accelerometers have an intrinsically noisy response (Subsection 2.1), the read response for the conditions depicted is not exactly 0 radians. Indeed, the original inclination received from the sensor has a static offset value (due ICs' misalignment regarding an assumed plain surface) and a complementary noise value. For this scenario, in order to assess the device's performance for each parameters combination's experiment, the static offset value was filtered through the subtraction of the collected samples' mean value from each sample, so that only the noise component is evidenced, as evinced in Figure 14. Each sample in the graphs of Fig.14 represents the noise complement of the sensing axis experiment's mean value.

Table 3 – Table portraying the combination of parameters that conducts the experiments from Figure 14

Experiment	Operation mode	Interval (ms)	Frequency	nValue	Timesleep (s)
Top	STANDARD	400	5	10	0.4
Middle	STANDARD	700	5	10	1.3
Bottom	STANDARD	1300	70	10	1.0

The Table 3 presents the combination of parameters for each horizontal couple of graphics from Figure 14. The inclination response read is given in milliradian. One may remark in the figure that each plot is scaled in the same y-axis boundaries, so that the discrepancy among

Figure 14 – Incliometer's response for three different parameter's settings.



Source: Author.

the noise response from each configuration of parameters can be comparable. Indeed, the stability of the device's response is extremely sensitive to subtle changes of the parameters' values, as it is clearly evidenced when comparing the response from each horizontal couple of graphics. For the sensing x-axis in the Figure 14, the difference of the maximum noise amplitude from the fig. top to fig. middle is 2.28 milliradian, and from fig. middle to fig. bottom is 0.35 milliradian. For the sensing y-axis, disregarding the outliers, the difference of the maximum noise amplitude from fig. top to fig. middle is 1.86 milliradian, and from fig. middle to fig. bottom is 0.096

milliradian. Such a huge difference of performance's behaviour can not be neglected for high accuracy applications.

It is known that noise components of any signal can be filtered by modern and sophisticated averaged-based algorithms. However a filtered response is an approximation of the real exact values. The less noisy is the signal response, the more accurate is the approximation of the averaged filter. For high accuracy applications, it is crucial the reduction of noise in the samples, once that noisy responses decrease the reliability of the signal's resolution.

The exposition of Figure 14 demonstrates the inclinometer's response for just three combinations of parameters, and in the presented situation the noise component responses distinguish remarkably from each other. Since each parameter can be settled within a wide range of values, there exist millions of possible combinations of parameters. To reliably assess each response experimentally is not a practicable option. A modern and feasible technique to achieve a high trusty intuition of the presented inclinometer's behaviour, for each combination of parameters, at a delimited configuration's volume, is the topic of this work.

Regarding the origin of the noise component in both the sensing axes of the inclinometer under analysis, a study carried out by Mohd-Yasin, Korman e Nagel (2003) and approached in Subsection 2.1 brings to light some evidences and answers of the typically noisy response of MEMS accelerometers. The authors concluded that intrinsic components of the sensing mechanism and the conditioning circuitry itself are responsible for the typical occurrence of noise in this technology. Since the foremost sources of the noise component are intrinsic to the MEMS technology, an enhancement of a stable state of operation is only achievable through the finding of an optimum adjustment of the device's parameters.

3.1.3 Inclinometer's application

As briefly mentioned formerly, the inclinometer device was developed in order to comply with the requisites of accuracy and versatility at the measurement of a machine tool structure's thermal distortion. By the measurement of the degree of displacement reported by several inclinometers throughout the machine tool, one becomes empowered to employ mathematical abstraction models, so that a map of the gradient of deformation of the machine is achievable. The researchers of the team of Large-Scale Metrology are interested to, at the time of writing of this work, state the transient error ($Err_{transient}$) of a machine tool, so that the volumetric error of operation at the tool center point might be fully characterised.

The volumetric error is defined as the sum of the static error (Err_{static}) and $Err_{transient}$. The error's static component is reasonably not laborious to define. In contrast, the transient component embraces the deformation of the machine's structure when performing movements. Depending on thermal conditions, load of operation, vibration and non-idealities of the environment, the transient component of the volumetric error is affected, and therefore the accuracy of the machine tool is impaired.

The inclinometer would so contribute providing data to a mathematical abstraction

model developed by the team, that a prediction of the transient error might be more accurate. In order to achieve such an aim, the noise reported by the sensors should not exceed 0.01 milliradian of amplitude, otherwise a recording of the machines' standard deformation would not be remarked. Therefore, the contribution of this work with the Large-Scale Metrology's goals is to provide an assessment of the inclinometer's potential by exploiting its combinations of parameters. In order to render the device more accurate, other techniques and calibration processes might be employed. However such themes are not under the scope of this work, once that this last is focused just on improving the sensor's measurement performance by characterising its behaviour for a given combination of parameters.

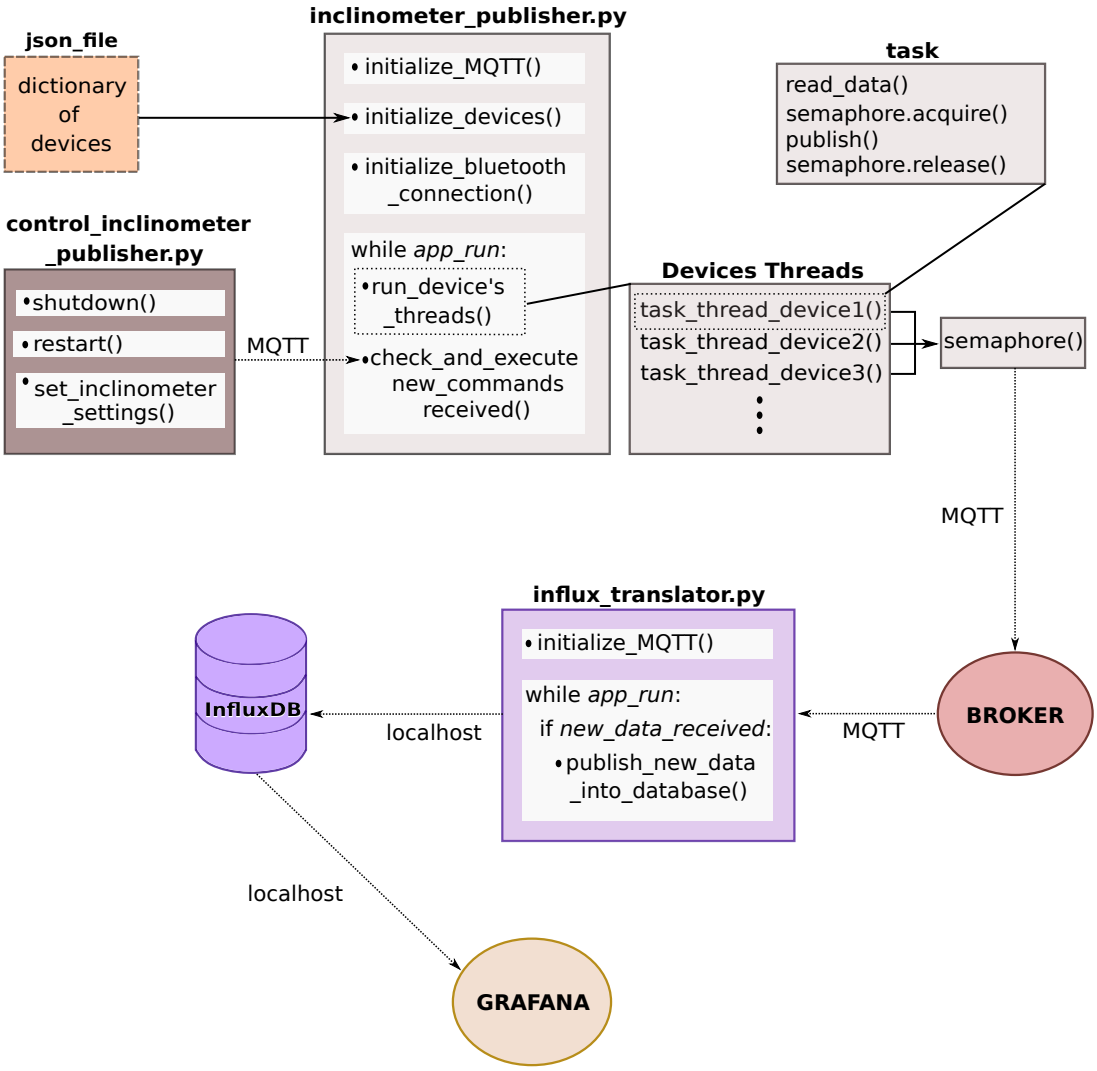
3.2 COMMUNICATION INFRASTRUCTURE

In order to carry out several experiments that assess the inclimoeter's performance, a dynamical software infrastructure was created by the author to accomplish the reading of several sensors at the same time, configure dynamically the parameters of these sensors and store safely and organized all the data read. The infrastructure was projected aiming robustness and reliability, so that one program could control another program in order to dictate the execution flow of instructions, and thereby, new parameters settings of the sensors could be settled automatically and at any time. Besides, was opted for the InfluxDB open-source database structure (INFLUXDATA, 2020) to carry out the storing data functionality, since databases are of easy manipulation and well organized for the handling with large amounts of data.

The communication among the programs in the infrastructure was accomplished through a standard messaging protocol for IoT, namely MQTT (2020). This protocol works through a publisher-subscriber paradigm, where the publisher module posts a message into a topic, and any module that is subscribed to such a topic receives the message. The management of the data flow among the topics is carried out through a program module known as Broker. In this work, the MQTT broker protocol employed to convey messages among the infrastructure was the open-source module developed by Light (2017).

The Figure 15 presents a schematic of the infrastructure developed. At first, the program **inclinometer_publisher.py** is executed in order to initialize its MQTT client module and to initialize the Bluetooth communication among the sensors specified in a file of type JSON. This program establishes two MQTT connections. The first, the program subscribes to a topic where is published commands that control its flow of execution. The second, for each new data read from some sensor, the program publishes this data read into a topic reserved to be subscribed by another program that is going to read this topic every time that new data is published. The **inclinometer_publisher.py** has two functionalities that are executed cyclically while the application is running: It manages the control of the threads that are going to establish the Bluetooth connection with each sensor and perform the transferring of data, and it checks whether there are new commands to be interpreted. These new commands can be either to shutdown/restart

Figure 15 – Schematic of the communication infrasctructure developed to carry out the reading of the sensors, the autonomous controlling of new settings of parameters and the storing of the data read into a database.



Source: Autor

the application, or to set a new combination of parameters for some specific sensor. In order to avoid the race condition problem (TANENBAUM; BOS, 2014), the publishing of new sensors' data by several threads into the same topic is controlled by a semaphore structure. The flow of execution commands sent to **inclinometer_publisher.py** is made by another program named **control_inclinometer_publisher.py**. In this program, one may configure a time-synchronized routine of commands to be sent. Through this functionality one may, for instance, let configured to be settled new combinations of parameters for some sensor at every 60 minutes, during a period of time of 8 hours. And hence, this setting of new parameters would be performed

automatically and within time constraints. The infrastructure is composed by another program, named **influx_translator.py**. This program is subscribed to the topic where is actively published data from the sensors. While the application is running, the program checks whether there is new data in the topic, and if yes, it publishes the data received into the InfluxDB database. Since the database structure is located in the same computer that runs the infrastructure, the communication between the **influx_translator.py** and the InfluxDB is made in the localhost. To finalize, once that there is any data in the database, this information can be visualized in present time with Grafana dashboard (GRAFANA, 2020). Grafana is an open-source web application multi-platform. This web application provides an efficient approach to graphically monitor time serial data stored in some database. By the means of graphically monitoring the inclinometer's data, it was possible to easily identify some patterns of the inclinometer's behaviour according to the change of its measurements parameters.

The robustness of the infrastructure developed is a crucial point for the automatically running of experiments, without error of execution, for hours. Indeed, in this work experiments were left running automatically for hours. Due to the functionality of controlling the inclinometer's parameters, it could be possible to configure the infrastructure to for each hour, a new configuration of parameters were settled, during periods of time up to 8 hours. The fact that all the thousands of data were organisationally stored in a database enabled reliability and practicality in the data treatment for the DOE analysis and machine learning dataset.

3.3 INCLINOMETER DEVICE'S SINGULARITIES

The inclinometer device has some particularities that characterize the response observed in the experiments of this work. In this section, one may find details of the inclinometer's behaviour and particularities of its measurement performance. The singularities here presented are treated and approached in the experiment's systematic performance (Section 3.4).

3.3.1 Outliers

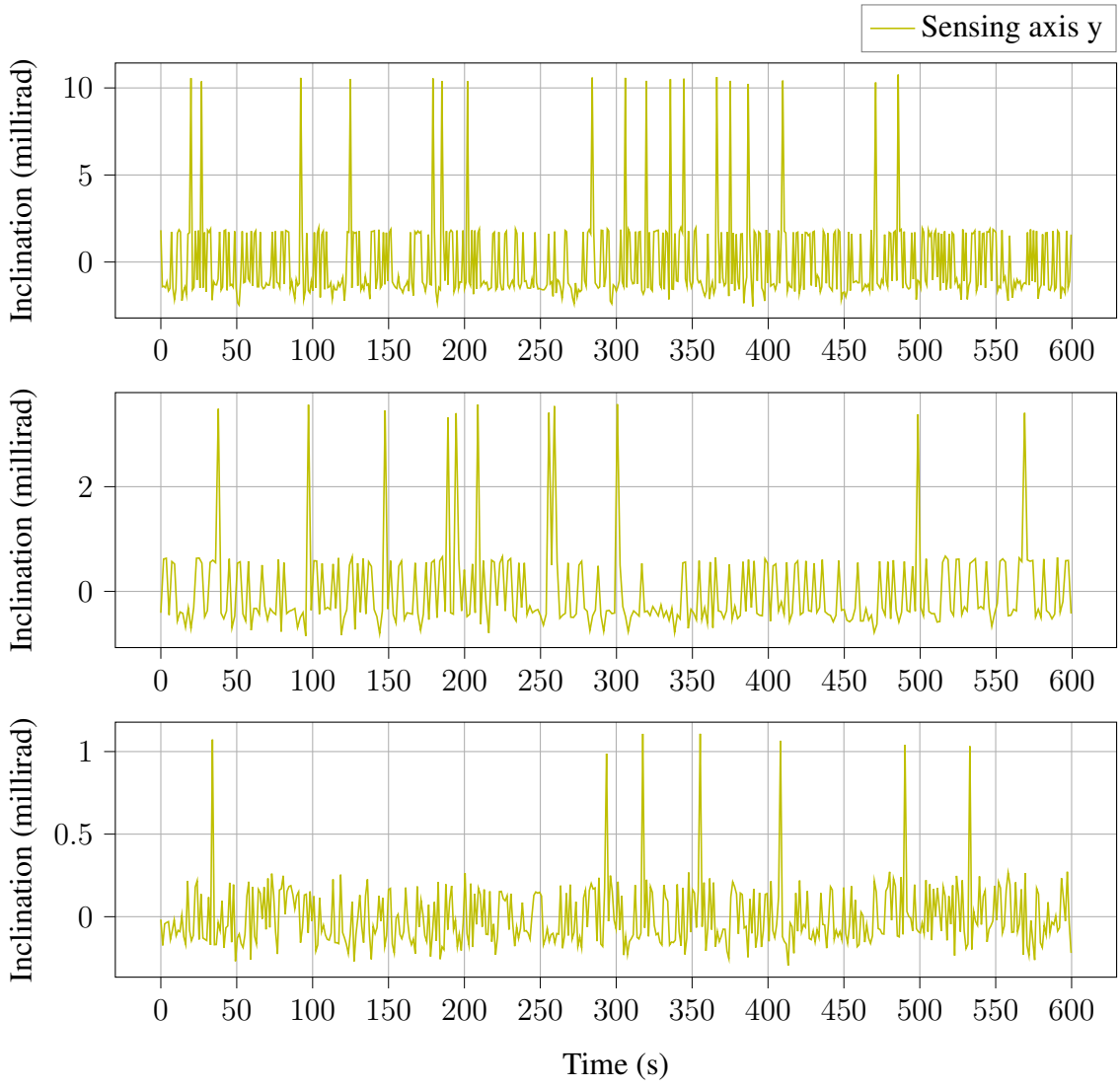
It is approached, in Subsection 3.1.2, the fluctuation of the sensor's response according to the settings established. A particular characteristic that also varies according to the combination of parameters is the frequency and intensity of outliers in the device's measurement. Before were defined the systematic experiment's approach, an immediate assessment of the inclinometer's response conceived the knowledge that for specific sets of parameters the presence of outliers is systematic and the characteristic of their amplitude and frequency are intrinsically related to the configured settings.

The Table 4 presents the settings of parameters of the experiments whose the response is found in Figure 16. The experiments of the Figure 16 were driven with the same methodology as the experiments from the Figure 12. The sensor was left laid still on a steady and assumed plain surface, reading its inclination for several minutes, so that the ideal measurement would report 0

Table 4 – Table portraying the combination of parameters that conducts the experiments from figure 16

Experiment	Operation mode	Interval (ms)	Frequency	nValue	Timesleep (s)
Top	STANDARD	500	5	10	0.4
Middle	STANDARD	1200	5	10	0.4
Bottom	STANDARD	500	120	10	0.4

Figure 16 – Inclinometer's response of the sensing y-axis for three different parameter's settings, where the occurrence of outliers is exposed.



Source: Author.

radians. The mean value of each signal received was respectively subtracted from the samples of each experiment, so that the graphics of Fig.16 present the noise component of the signal. In the figure, is presented just the response for the sensing y-axis, although the outliers are also found

in the measurements from the sensing x-axis. One may remark that the three graphics are in different y-axis boundaries, so that the amplitude of the outliers can be more easily comparable with the noise's amplitude from that specific set of parameters. In Fig.16 top, the frequency of outlier samples is quite high, and the noise component reaches the amplitude of approximately 2 milliradians, whereas the outliers' amplitude reaches the value of 10 milliradians. In Fig.16 middle, yet is found a high frequency of outlier samples, but it is fewer than the fig. top. The noise's amplitude rounds the value of 0.6 milliradian, whereas its respective outliers' amplitude measured is 3.4 milliradians. And at Fig. 16 bottom, the amount of outliers is visible much lower than the other two configurations depicted. The noise component reaches the amplitude of value 0.27 milliradians, whereas its outliers' amplitude rounds the value of 1 milliradian.

The presence of variable-value outliers is a fact that can not be untreated in the strategies further presented. Outliers can totally destroy statistical analysis and hence conveying wrong assumptions of the mean value of some signal. In another hand, since the frequency of outliers is also variable, the presence of this phenomenon may indicate that such combination of parameters leads to an unstable measuring condition. Thus, future analyses must treat the intensity and frequency of outliers, so that statistical evaluations are not ruined and yet such instability is pondered.

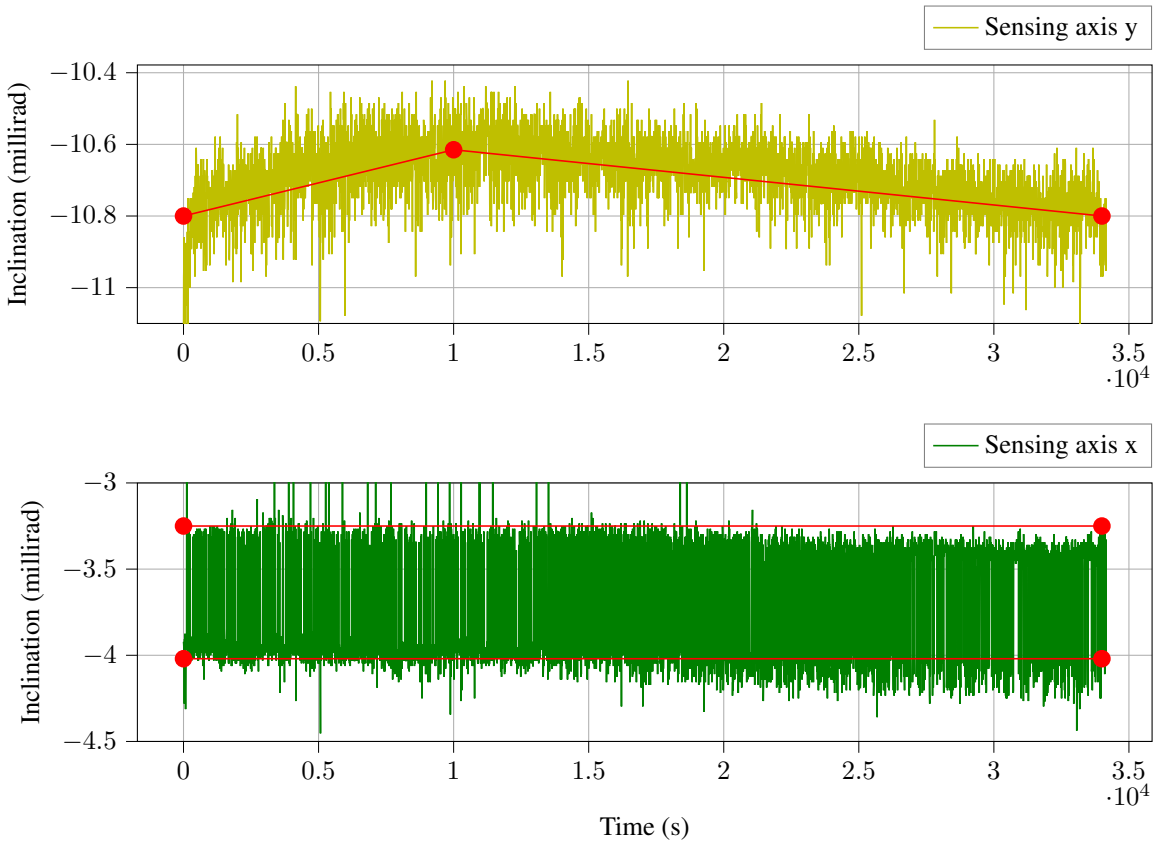
3.3.2 Bending effect

In order to assess the stability of the inclinometer device for long periods of unrelenting measurements, experiments were so accomplished, that the sensor was left laid still on an assumed plain surface collecting data for approximately 10 hours. These set of long period experiments were carried out with two inclinometer devices for three different configurations of parameters. For this subsection, it is approached the raw measurement of just one of these experiments and just for one device, as presented in Figure 17. More details about these long period experiments comparing the performance of two inclinometers devices are approached in Appendice B.4.

Concerning the long period experiment of Figure 17, some important characteristics may be observed and analysed in the inclinometer's response for its two sensing axis. One may visually remark in the figure that the noise component of the inclinometer's measurement remains at the same pattern, i.e. has roughly the same performance, during the whole period of 10 hours of measuring. Further it will be presented a metric to assess the noise component of an experiment, and this very remark for the Fig. 17 will be approached with more details in Section 3.4. Another important aspect that can be visually remarked is that the tilt measured during the 10 hours does not remain constant. Its performance for both the sensing axis shows apparently to slowly bend through the hours of measurement. Since the sensor device was laid still on an assumed plain surface, and this surface is not supposed to bend at any time, this bending effect in the measurements conveyed by Figure 17 is not expected.

As it was formerly mentioned, these long-period experiments were made with two

Figure 17 – Inclinometer's response for both the sensing x-axis and y-axis for a long period experiment, where is evidenced a bending effect in the measurements.



Source: Author.

devices and with different sets of parameters. The results for all these experiments revealed the same bending effect in the measurement response. However, the bending effect of each experiment has its own shape performance. Each time that a long period experiment was made, the bending curve assumed a different shape aspect. For the response of the experiment of Figure 17, the bending of the sensing y-axis is given upward for the first 10000 seconds, and afterwards its direction is changed to downward. For the sensing x-axis, the measurement remains unbent for the first 17000 seconds, and afterwards a bend effect begins downward.

In order to understand this issue, some hypothesis are risen. It is possible that, since the inclinometer devices are manufactured by hand, imperfections of the soldering may lead to noise aspects into the circuit that communicates with the accelerometer modules. This noise could destabilize the reading of acceleration inside the accelerometer modules. Another explanation for the bending effect, and this is more plausible, is that the inclinometer is under environment variations. As approached in Subsection 3.4, all the experiments were carried out in the same place, such place that is located in the 7^{th} floor of a building. It is quite reasonable that the tower may slightly bend due to weather conditions, and the effect of this slightly bending would be felt by high sensible accelerometers located at the 7^{th} floor of the building. The

question of environment interference could be ascertained by assessing experiments made in other location. However this work is written in the period of the COVID-19 pandemic, and long period experiments are not feasible in other locations at the moment of the writing of this work.

Although the source of the bending effect is not certainly defined, still there is a need to treat this aspect, so that the bending effect does not impair the DOE and machine learning approaches to assess a reliable characterisation of the inclinometer device. Strategies to handle with this issue are presented in Section 3.4.

3.4 THE EXPERIMENT'S SYSTEMATIC PERFORMANCE

In order to validate all the experiments carried out in this work, a systematic methodology that drives and processes the experiments had to be developed, so that every experiment could be conducted and analysed under the more similar conditions as possible.

In this work, is defined as an experiment the procedure of configuring the inclinometer device developed by the team of Large-Scale Metrology with a set of parameters, leave this device laid still in an assumed plain surface collecting data for several minutes, so that the ideal result of all the measurements were expected to point 0g (also 0 milliradian when applied the eq.9). But since the device is not ideal, the measurements have a static offset (probably resulted from misalignment effects due to handmade manufacturing), outliers and a noise component. Since the misalignment effect is not a concern of this work, the performance of the inclinometer for the situation described is assessed through the evaluation of the noise component and outliers, so that as a final output of the processing of an experiment, one single real-value that represents the level of noise and outliers of the experiment is provided for the sensing x-axis, and another for the sensing y-axis.

All the experiments of this work were performed through the infrastructure described in the Subsection 3.2, at the same place and the same period of time of the day. The place that the inclinometer device remained still collecting data is located on the surface of a commode, inside a bedroom that is found in the 7^{th} floor of a student building. This location for experiments with high sensitive sensors is not ideal, but since the experiments of this work were carried out during a lockdown unleashed by the COVID-19 pandemic, other locations were not possible to be employed at the moment. Due to the fact that a student building is not quite still during the daytime period, all the experiments were carried out in the late of the night until the dawn. To be precise, the experiments were carried out in between the range from 21h to 8h. Since the settlement and configuration of a new combination of parameters is autonomously done by the infrastructure developed, several experiments could have been driven automatically in one night, while the most people of the building, including the author, were sleeping, taking advantage thereby of the stillness of the building, so that the experiments could endure the minimum of environment's interference as possible.

In this work, most of the experiments lasted about 60 minutes. Was common the

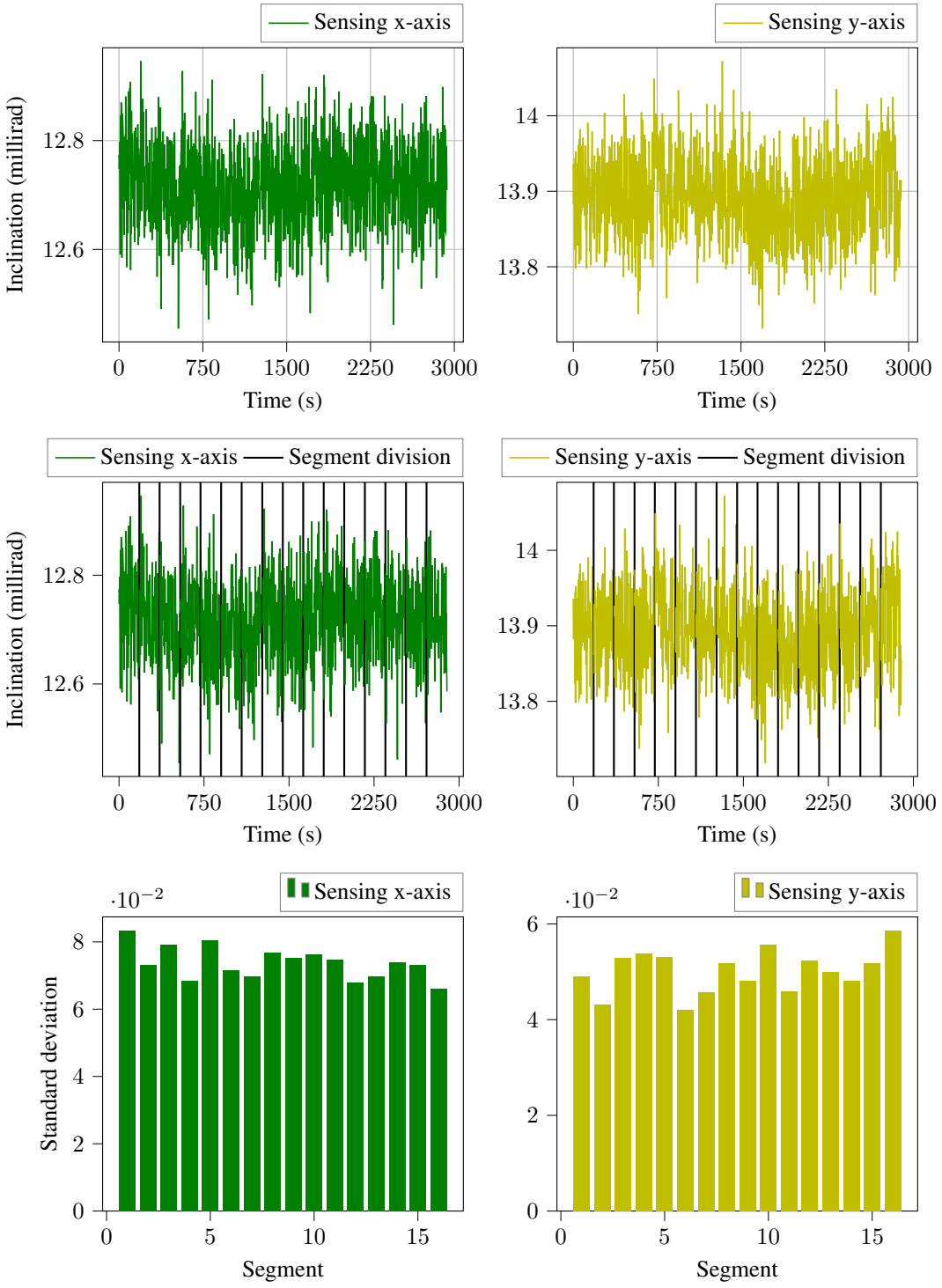
execution of eight experiments with eight different combinations of parameters in one night of autonomous experiments. However, in order to avoid some systematic bias concerning the constant change of the devices' parameters, a protocol that sets a new parameters' setting was developed and deployed. Each time that a new setting was configured to be established, the inclinometer device was set to stop the measuring, the Bluetooth connection was broken and the device remained inactive for five minutes. After this period, a new Bluetooth connection was established, and a new combination of parameters was configured for the next performance of the measurements. This protocol was followed by every experiment in this work, so that if in one night eight experiments were carried out, eight times this protocol would be executed.

Every experiment's measurement data was submitted to the exact same data processing. The Figure 18 and 19 summarizes graphically the main steps of this processing. After an experiment is accomplished, its samples are queried from the database, and the samples belonging to the first six minutes of the experiment are disregarded. This disregarding strategy is so done that possible instabilities due to the stabilizing time of a new combination of parameters of an experiment does not interfere in the evaluation. The result of this operation is shown in Figure 18 top, for both sensing axes. Following, the entire experiment's data is temporally segregated into segments of three minutes, so that each segment has sequentially in timeline the inclination data belonging to three minutes of duration, exactly how is depicted in Figure 18 middle, individually for each the sensing axis. This last observation means that the operation of segmentation is performed regarding individually the time of inclination reading of each sensing axis, so that each sensing axis owns its own timeline.

Each segment from each sensing axis is submitted to an outlier filter based on median values. This filter works as following for each segment: First, is taken the median value of that segment. Then, is computed the module of the difference of each sample from the segment with the median value taken. At this stage is found a vector of differences. Subsequently, each difference from the vector of differences is divided by the median value of the very vector of differences, and the result of this operation gives a value s for each sample in the segment. Then, is defined a threshold number, so that is defined as an outlier sample, if a sample whose s value is equal or greater of this threshold. The outlier sample is then deleted from the segment. In this work, the threshold number used was the same for all the experiments, and it was determined empirically, namely the value 80. The considerations to determine this threshold were that just extreme huge outliers values, that would destroy the following analyses, were deleted. Minor outliers values, that still can slightly penalize the following analyses were maintained. The decision to do not filter minor outliers is taken in order to consider the high frequency of their occurrence as a signal of instability, consequent of the combination of parameters configured.

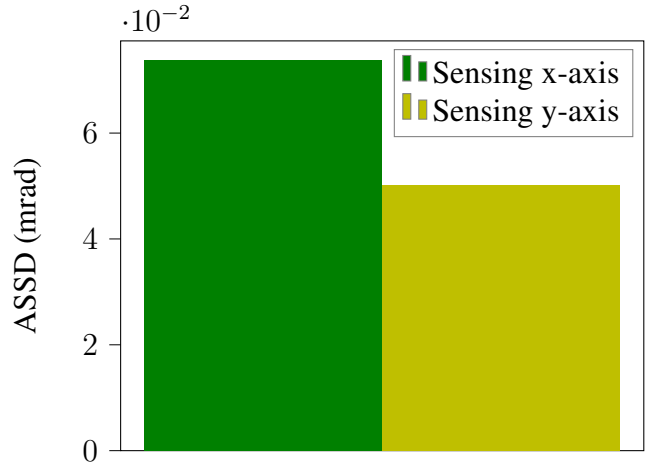
After the outlier filtering operation, is computed for each segment the standard deviation of the measurements with Bessel's correction, given by the Equation 18, where s_{seg} is the standard deviation of the samples from a segment, M_{seg} is the total amount of samples in that segment, t_i is the i^{th} sample of a segment and $\bar{t}$ is the mean value of the samples contained in

Figure 18 – The figure presents graphically some steps of the experiment systematic performance. In fig. top, is presented the signal after the disregarding of the samples belonged to the first six minutes of experiment. Fig. middle presents the segmentation of the signal from fig. top. And in fig. bottom is found the standard deviation related to the samples of each segment.



Source: Author.

Figure 19 – The figure presents the value of the Average of the Segment s' Standard Deviation from Figure 18.



Source: Author.

a segment. The Figure 18 bottom presents the representation of the standard deviation of the tilt measurement from each segment, for each sensing axis. One may remark that the standard deviation metric is sensible to outliers, and thus the well regularization of the outlier filter is crucial to employ this metric into segments.

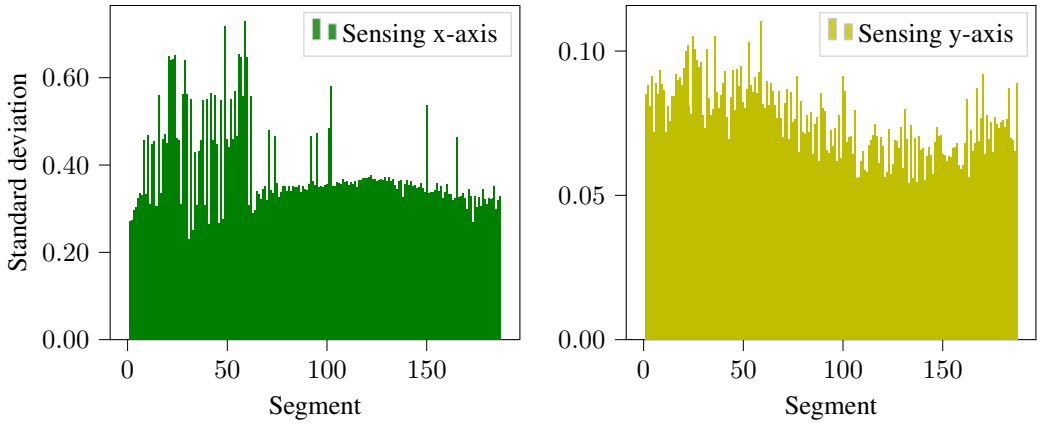
$$s_{seg} = \sqrt{\frac{\sum_{i=1}^{M_{seg}} (t_i - \bar{t})^2}{M_{seg} - 1}} \tag{18}$$

The final step of this methodology is to average all the segments' standard deviations of one sensing axis, in order to obtain a single value that can entirely express the experiment's performance. The Figure 19 presents this final step, for the experiment depicted in Figure 18. This single value expresses the level of spreading of the tilt's measurement from its mean value. In this work, this metric is employed to assess the performance of the inclinometer, so that the closer to 0 is the value of the Average of Segments' Standard Deviation (ASSD), the smaller is the noise component in the measurements driven by the parameters' combination settled, and thus such set leads to a more stable state of operation of the device.

The approach of to temporally segment the experiment was developed in order to dissolve the environment's interference in the ASSD. Besides, such approach also minimizes the bending effect's issue, since that the bending occurs through the hours, and the segmentation is temporally applied in the scale of minutes, so that the bending effect is not strongly felt in the scale of a segment, and thus the ASSD metric is not severally penalized by this interference in the measurement's performance.

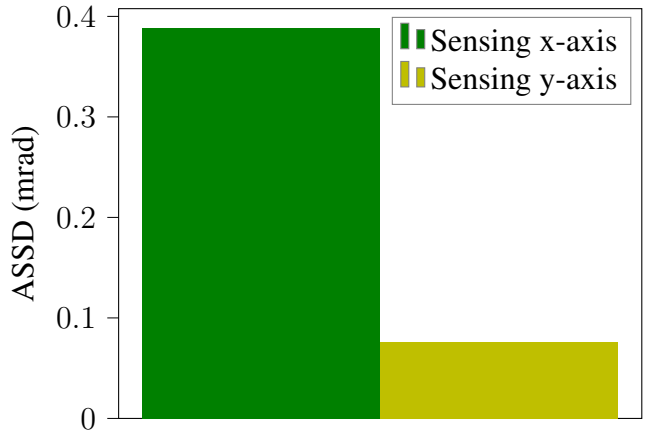
The Figure 20 portrays the standard deviation of the samples from the long time experiment of Figure 17. In Figure 17 it is evidenced the bending effect. One may remark in

Figure 20 – Standard deviation of each segment from the experiment from Figure 17.



Source: Author.

Figure 21 – The figure presents the value of the average of the segment s' standard deviation from Figure 20.



Source: Author.

Figure 20 the effect of minor outliers in the standard deviation of some segments from both the sensing axis. Each of these numerous segments contributes hence with information of its samples' behaviour into a uniform average, as conveys the Figure 21.

From Figure 20, is possible to observe that during a period of almost 10h, the standard deviation of the samples from the segments underwent small variances. Since the scale of their standard deviation remained the same, the averaging of their segments' standard deviation must provide a satisfactorily reliable numerical value that one may employ to numerically assess the performance of an experiment.

3.5 INCLINOMETER'S DOE MODEL

As demonstrated so far, the inclinometer device assumes different performances for different sets of combinations of parameters. The goal of this work is to present the deployment of the hybrid ML/DOE technique employed to characterise the features of the smart inclinometer device. In order to accomplish such aim, it is presented in the Appendice B the deployed DOE methodologies. Since the DOE strategies and experiments carried out in this work require a detailed analysis and inspection, in order to do render a smooth reading structure, the results and approaches concerning the inclinometer's DOE are placed in the appendice.

Some decisive conclusions are remarked with the inclinometer's DOE. Concerning the ASSD metric developed to assess the performance of the sensor, it became proved stable and reliable. Regarding the behaviour of the performance due to the parameter's set, the interactions among the parameters' combinations of levels were classified as mostly nonlinear, fact that renders challenging the visual analysis of the set of parameters' pattern. It was also observed that the set of parameters can be classified into two groups, each group having its own particularity regarding the variation of the ASSD score. It has been attested that when a set of parameters leads to an ASSD score greater than the rank of 0.1, this very achieved ASSD value has a variability notably higher than ASSD scores below the rank of 0.1, reached by another set of parameters. Performances that have achieved ASSD values below 0.1 have a trend to reach almost the same ASSD value, repeatedly. In contrast, performances that have led to ASSD values greater than 0.1 have a trend to diverge from former ASSD values reached by the same conditions. Such is the discrepancy of ASSD scores' steadiness between these two categories, that new strategies may be developed in order to analyse each group's behaviour separately. Another important characteristic remarked at the DOE model is that each sensing axis of the sensor has its own particularities, so that a set of parameters leads to a different quality of measurement for each sensing axis.

The DOE model provides insights of the sensor's behaviour, so that one might estimate the performance of the device for a yet untested set of parameters. However, by the reasons commented, this is not a trivial endeavour, and mainly due to the non-linearity among the interactions of parameters, not accurate. The DOE model lacks of a decisive tool that can state, with a margin of accuracy, the performance of the inclinometer for any set of parameters, already tested or not. Based on the conclusions achieved by the DOE model, the machine learning model is proposed in order to complement the observations so far remarked. The a priory information acquired by the DOE can refine the machine learning strategies and algorithms, so that an accurate personalised machine learning model for the inclinometer can be accomplished. For instance, since one can classify the ASSD values into two groups, each group can be evaluated by a metric that best can assess the particularities of those data. In Section 3.6, the inclinometer's DOE conclusions are explored in order to provide theoretical foundations on which the machine learning approach substantiates itself to become accomplished. The results and main aspects of

such DOE are recapped whenever mentioned.

3.6 INCLINOMETER'S MACHINE LEARNING MODEL

In the section 3.5 is approached a strategy to characterise the inclinometer's behaviour for several combinations of parameters, by the means of experiments and evaluations driven through the DOE methodology. In this section, is presented the employment of modern machine learning techniques to complement the analyses achieved through the DOE methodology.

The conceptualization of ML technique here presented relies, in order to be driven, on former assessments of the inclinometer's features. The DOE analyses provide some conclusions that are here explored. Immediately, the result of the several experiments performed with different parameters' values and operation modes conveys the knowledge that the device's performance is affected by the set of configurations that dictates the sensor's measurement routine. Had the inclinometer no performance's variation due to different configurations of its measurement routine, no necessity of a characterisation would exist. Nevertheless, the DOE evaluation exposes clearly that the performance's accuracy and variability of the acceleration measured is affected by a logical reasoning among the parameters' values. It became evident that certain patterns of configuration lead to more a stable, reliable and accurate performances than others. The design of experiments could also attest the reliability of the ASSD metric to assess the inclinometer's performance, so that the metric provides a real-valued number that assimilates the goodness of performance. The device's responses were graphically explored through a delimited volume of parameters, by the deployment of cube and square plots. These graphic tools can convey a good notion of how to determine the best settings for a specific application that requires a specific set of parameters. However, when there exists non-linearity relationship among the parameters, the graphic DOE tools can not provide a satisfactory knowledge of behaviours of responses for settings inner the volume of parameters explored, as it is the situation of the inclinometer device. The interaction plots in the Appendice B, for the standard and repeated operation mode, portrays the situation where the gradient of better performance changes its direction as inner steps of experiments are carried out. Since the changing of parameters' interactions is not predictable by the DOE approach, there is no assurance that a chosen point inside the parameters' volume will follow the patterns of response so far remarked. Hence, the evaluation of configurations of parameters inner a delimited volume of parameters becomes a considerably complex task, once that the parameters' relationship is not fully characterised, but rather outlined.

The machine learning technique consists of relying on a dataset with maximum significance, provided by a DOE design, in order to train a predictive regression model (Section 2.4) built to predict, with satisfactory accuracy, the behaviour of the inclinometer device for every possible combination of parameters constrained into a chosen delimited volume of parameters. The employment of this strategy is enabled by the conclusions remarked in the DOE approach. Once that the DOE methodology could attest the reliability of the ASSD metric, and could

enlighten the patterns of the ASSD's variability for different sets of parameters, the machine learning approach can be carried out discerningly, and decisions of project can be taken relying on previous observations. For instance, it is known that the sensor's responses do not follow a linear reasoning according to the given parameters' values. Thus, it is more probable that better predictions can be achieved through the training of regression models that has a non-linear learning basement. By the DOE approach is also known that certain levels of ASSD value has higher variability than others. This knowledge can be explored to create a customized metric to assess trained ML models.

In this section is presented the methodology of the machine learning approach, relying on the DOE observations, developed and deployed in this work in order to complement the characterisation outlined by the DOE methodology. First, is presented the structure and procedure to acquire a dataset with maximum significance whose the ML models are trained. Further is presented the metrics employed in the ML approach to assess and judge the performance of experimented algorithms. Subsequently is found the design of the ML methodology and the regression models chosen to accomplish the predictions. Afterwards, the process of the models' training and tunning is described. Concluding the section, is presented how the models so far built are validated.

3.6.1 Inclinometer's Dataset

The inclinometer's machine learning approach is developed in order to predict the ASSD values for an unobserved configuration of parameters inside a delimited volume of parameters. Aiming to accomplish this goal, two datasets were collected by a single inclinometer device. One to characterize the standard operation mode, and other to characterize the repeated operation mode. The standard operation mode's dataset is consisted of three independent variables, namely the operation mode's DOE factors: *interval*, *frequency* and *timesleep*; and two dependent variables, which are the respective ASSD values of the sensing axis x and y. In turn, the repeated operation mode's dataset is consisted of two independent variables, namely *interval* and *frequency*, and two dependent variables which stand for the respective ASSD values of the sensing axis x and y. Since the continuous operation mode is already totally characterized by the DOE approach, as detailed in the Appendice B, the machine learning approach is not applicable to it.

The standard mode's dataset is structured as train and test set (Subsection 2.4.1). The dataset was planned to comply with a DOE design, so that the train set defines a delimited volume of parameters, and the test set consists of unobserved responses inside this delimited volume. The choice of combinations of parameters that consists the train set was so defined that the volume of parameters complies with a 3^4 full factorial design of experiments. The factors and levels of the standard mode's train set design are presented in Table 5. The levels of the factors were so chosen that the volume of parameters is filled with observations roughly equally distanced in the volume's space.

Table 5 – Table presenting the factors and factors' levels of the standard operation mode's train set design, which is consisted of a 3^4 full factorial design.

Interval	Frequency	Timesleep	nValue
400	5	0.4	10
700	30	0.7	10
1000	70	1.0	10
1300	120	1.3	10

Table 6 – Table presenting the factors and factors' levels of the standard operation mode's test set design, which is consisted of a 3^3 full factorial design.

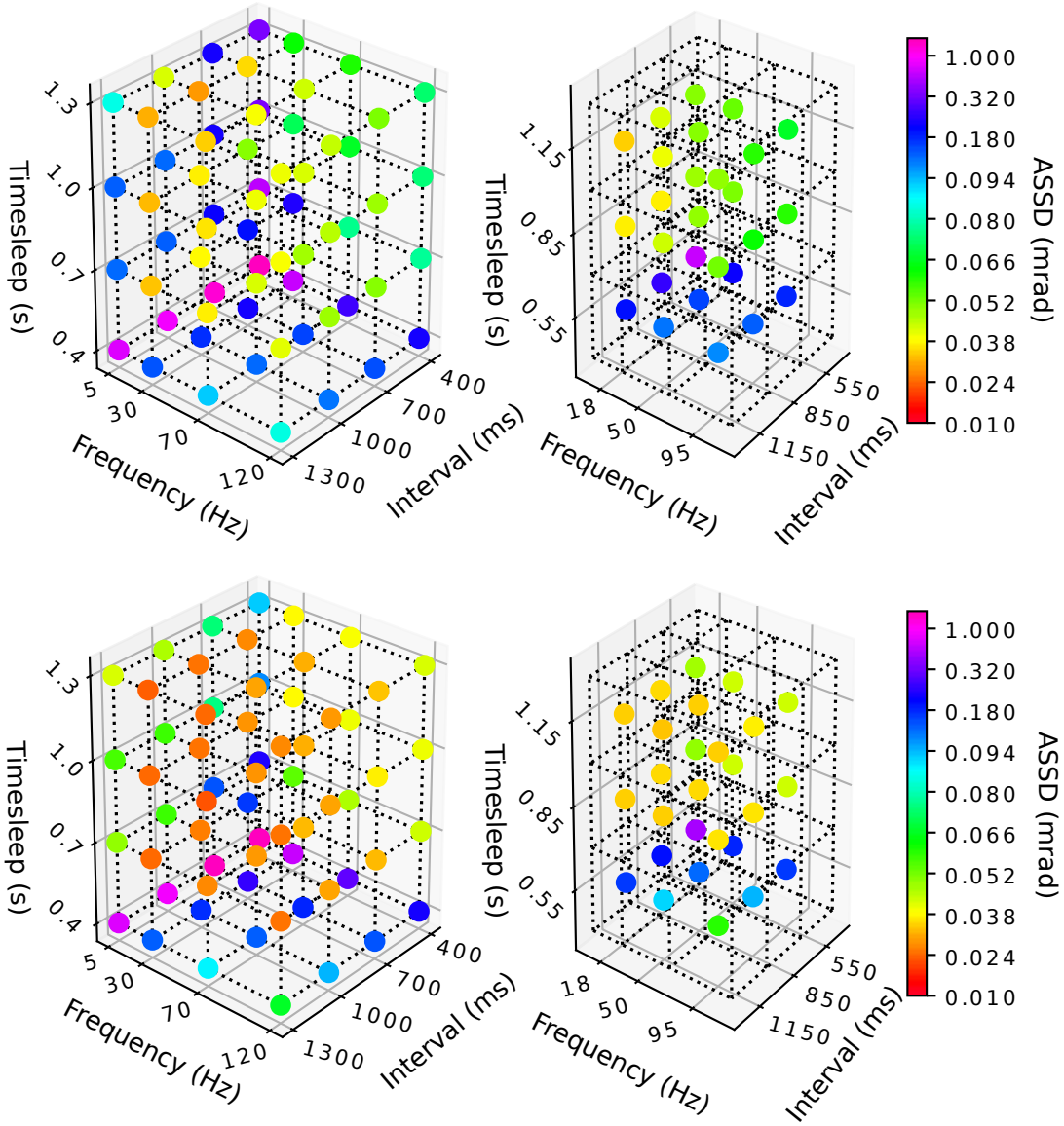
Interval	Frequency	Timesleep	nValue
550	18	0.55	10
850	50	0.85	10
1150	95	1.15	10

The standard mode's test set structure was also defined following a full factorial design, but with just three parameter's levels, so that the design is classified as a 3^3 architecture. The levels' values and factors regarding the standard mode's test set are found in Table 6. The levels' values were so defined in order to remark a response that is equally distanced, in the volume of parameters, of the eight nearest points belonging to the train set.

The Table 34 and Table 35 in the appendice C present the exactly train and test set which were deployed in the building of the regression models for the predictions regarding the standard operation mode (disregarding the columns of operation mode, trial and nValue). The data of both the operation mode's dataset were collected as dictates the systematic methodology (Section 3.4), but without considering the randomization of the trials. Besides, each train set was collected once, and thus, the data employed to train the ML model are not an average of several runs of trials, as they are in the DOE characterization model. The conclusions provided by the DOE approach, detailed in the Appendice B, lead the assumption that the ASSD value of one performance, driven by one set of parameter, does not reach enough variability to report a different pattern from the reported by the average of ASSD response achieved by several performances of the same set of parameter. Even though, the acquirement of the standard train set was carried out just once, yet it was required eight nights of experiments (eight experiments per night) to fulfil the designed volume of parameters. The standard mode's test set was collected three months after the acquirement of the standard mode's train set, and required three days of experiments (nine experiments per night) to fulfil the designed volume of parameters.

The Figure 22 portrays the standard operation mode's dataset as cube graphics representations. The column of cubes on the left side stands for the train set, whereas the

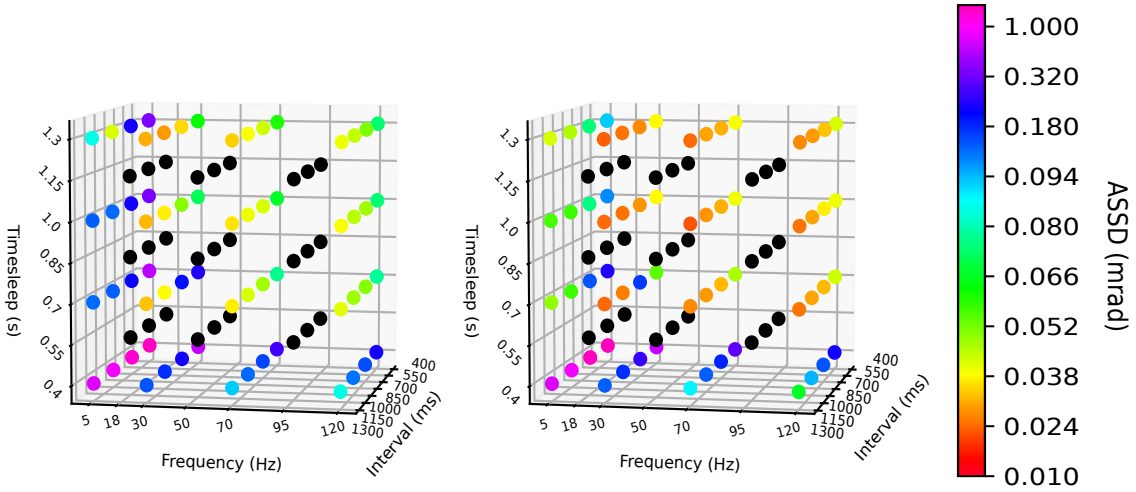
Figure 22 – Standard operation mode's dataset. The column of figures in the left stands for the train set, whereas the column in the right stands for the test set. The first row of figures portrays the data regarding the sensing x-axis, whereas the second row refers to the sensing y-axis' data.



Source: Author.

cubes on the right side stand for the test set. The first line of cubes refers to the sensing x-axis' responses, and the sensing y-axis' responses can be found in the second line of cube plots. Through the picture, one may remark the standard mode's delimited volume of parameters that bounds the standard mode's dataset. The train set's delimited volume conveys the patterns of the inclinometer's ASSD responses for the given operation mode. In fact, through the train set presented, some impressions of the devices' behaviour can be visually, and considerably well, observed. The characterization of the full volume of parameters, through a regression model, is

Figure 23 – Standard operation's mode dataset. The figure evinces a delimited volume of parameters, which stands for the train set, presented as scaled colored points, in contrast with the test set, located inward the delimited volume, presented as black points. The cube plot placed in the figure's left side stands for the sensing x-axis' responses, whereas the right side stands for the sensing y-axis' responses.



Source: Author.

expected to be accomplished by the means of the model's remarking of the responses' pattern expressed by the high significant train set provided. The performed predictions are to be assessed attesting the similarity with the test set presented in the Figure's 22 right column of cube plots. As formerly commented, the test set data were arranged to be so placed inside the volume of parameters, that each test set point is equally distanced of the eight nearest points belonging to the train set. As one may visually observe in Figure 22, each test set point is found in the centroid of a cube whose the vertices are train set points. Such configuration enables one to evaluate the goodness of remarking patterns of the built regression model, once that the algorithm is going to be validated by the predictions' performance of the points most different as possible of the points that had trained the regression model. The Figure 23 evinces the placement of the test set, represented as black points, inside the train set's delimited volume, presented as scaled coloured points. Figure 23 left stands for the sensing x-axis' responses, whereas figure 23 right stands for the responses of the sensing y-axis.

The repeated mode's dataset was designed and structured following the same systematic of the standard mode's dataset. The repeated train set was so architected to comply with a DOE design, that a delimited volume of parameters is defined to provide significant observations of the inclinometers' performance under the repeated operation mode. The choice of parameters' values of the train set was structured to comply with a 2^7 full factorial design, and the factor's levels are presented in Table 7.

The repeated mode's test set was also structured in a DOE architecture and can be classified as a 2^6 full factorial design. The repeated mode's test set values were so defined that

Table 7 – Table presenting the factors and factors' levels of the repeated operation mode's train set design, which is consisted of a 2^7 full factorial design.

Interval	Frequency	nValue
300	7	10
500	20	10
700	40	10
900	60	10
1100	80	10
1300	100	10
1500	120	10

Table 8 – Table presenting the factors and factors' levels of the repeated operation mode's test set design, which is consisted of a 2^6 full factorial design.

Interval	Frequency	nValue
400	13	10
600	30	10
800	50	10
1000	70	10
1200	90	10
1400	110	10

each point in the test set is equally distanced of the nearest four points in the train set delimited volume of parameters. The Table 8 present the factor's levels of this operation mode's test set.

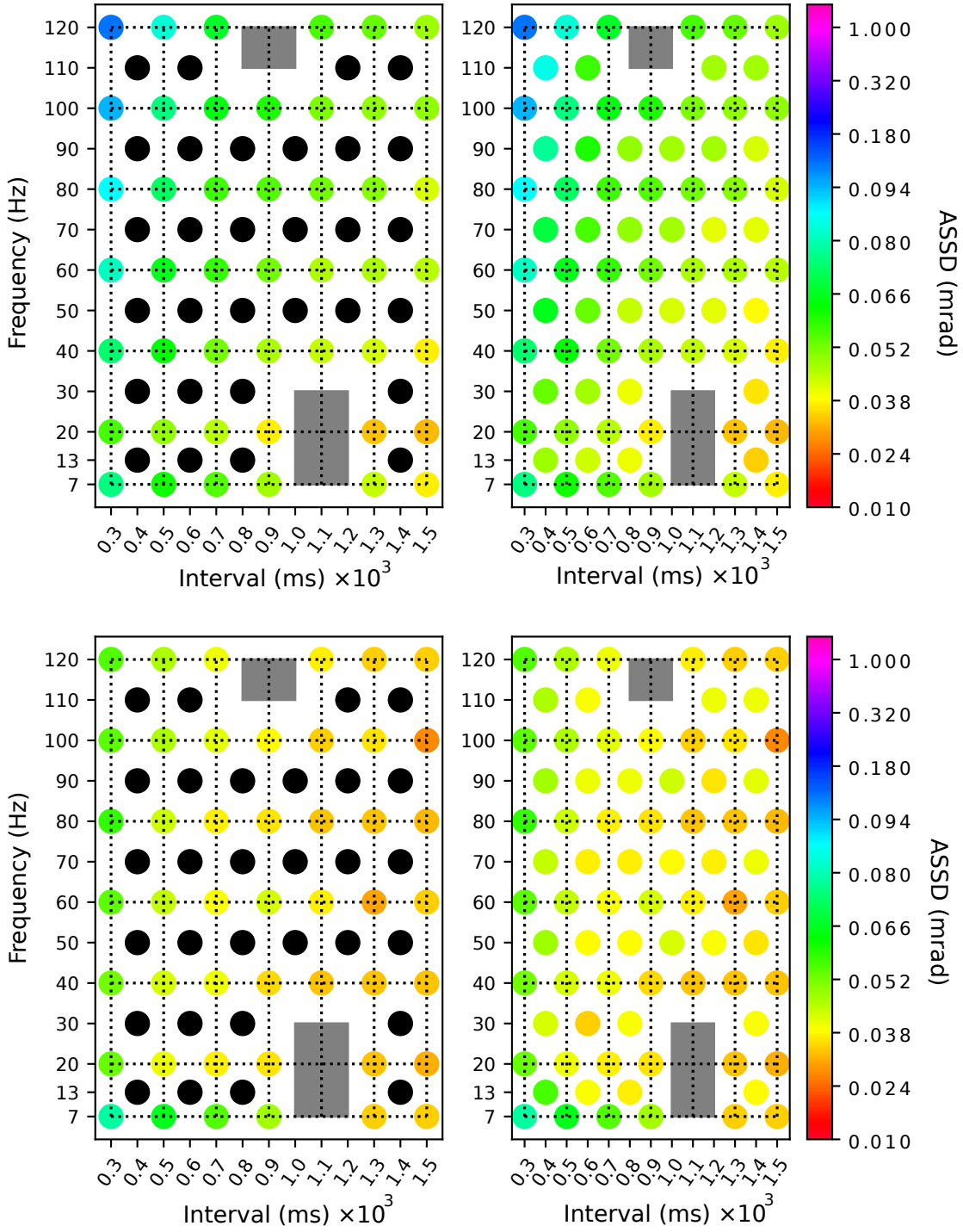
The tables 36 and 37 in the appendice D present the repeated mode's train and test set, respectively, that were deployed in this work to train and attest the regressions models regarding the repeated operation mode's characterization (ALso disregarding the columns of operation mode, trial and nValue). As it is for the standard mode's data, the repeated train and test set were acquired and processed through the exact same systematic methodology described in Section 3.4, save for the randomization of the trials, which was not applied in the machine learning data acquirement. And as carried out for the standard mode's train set, the repeated mode's train set was collected just once, so that the final ASSD values employed to train the ML model are not an average of several runs of trials, as they are in the DOE approach. The DOE characterization of the repeated mode had led to the observation that the variability of the repeated mode's trials is not considerable to depend upon an average, in order to remark the patterns reported by the operation mode's performance. The ASSD response of one performance, driven by a set of parameters, is not variably enough to report a value out of the pattern reported by the average of ASSD responses of several performances with the same set of parameters. Nevertheless, the

acquirement of the repeated mode's train set required six nights of experiments (roughly eight experiments per night) to fulfil the parameters' volume designed. The respective test set was collected four months after the train set's acquisition, and required three nights to be fulfilled.

Figure 24 portrays the repeated mode's dataset. The left column of graphics stands for the train set, depicted as scaled coloured points, in contrast with the test set, as black points. The right column of graphics stands for the train and test set, depicted both as scaled coloured points. The first row of graphics represents the responses of the sensing x-axis, whereas the second row stands for the sensing y-axis' responses. Through the figure's left column, one may visually remark the architecture of the DOE approach into the dataset structure. The train set bounds a volume of parameters that significantly contemplates the device's performance for the repeated operation mode. The test set is so placed that each point is located at the centroid of a delimited square whose the vertices are the nearest train set points. Such a configuration enables the built regression model to be tested to predict points most different as possible of those whose the model had been trained with. The grey squares on the graphics in Figure 24 represent configurations of parameters that could not have been settled. For the device whose the experiments concerning the machine learning approach were performed, whenever the parameters were settled in such a configuration inside the bounds of the grey squares, the Bluetooth communication reported complications at the fulfilment of the measurement and data broadcasting. This singularity was not ascertained, and the reasons for such behaviour remain, at the time of this writing, unknown. A curious observation remains due to the fact that the performance with combination of parameters consisted of *Frequency* equal to 120 and *Interval* equal to 900 could have been reached in the DOE experiments, in Subsubsection B.3.2, but was not achievable for the experiments concerning the machine learning approach, although the procedure carried out to acquire an ASSD value was the same for the DOE and ML experiments. The regions inner the repeated mode parameters' volume that may lead to a nonfunctional state are bypassed in the train and test set.

Both the datasets so far presented rely on former knowledge conveyed by the DOE approach to be designed. Further, the variability analysis leads the observation that high ASSD values, in the standard mode, have a trend to present significant ASSD's variability. And since the standard mode's dataset acquired is composed by some high ASSD values, the validation of the regression model must consider the variability effect in the assessment methodology. The datasets collected were so designed to feasibly provide data to train regression models, in order to reliably predict the inclinometer's performance, assessed through the ASSD metric. Since the methodology developed to evaluate the sensor's performance requires a considerable amount of time to acquire just one single vector of independent and dependent variable, the design of the standard and repeated mode's dataset could not have a plenty amount of data. It is known that the more the amount of significant data a dataset has, the more accurate and generic the built ML model will be (subsection 2.4.1). Yet, both the datasets acquired have a scarce amount of data. The standard mode's train set is composed by 64 observations, whereas the repeated's mode

Figure 24 – Repeated operation's mode dataset. The left column of graphics stands for the train set, as scaled coloured points, and as black points is represented the test set. The right column of graphics represents the train and test set, both as scaled coloured points. The first line of graphics regards to the sensing x-axis' responses, whereas the second line concerns to the sensing y-axis' responses.



Source: Author.

train set is composed by 46 observations. Such dataset's size limitation could be improved by

acquisition of more data, but such an accomplishment is not feasible when the resources for the data acquirement are limited or too expensive. For the inclinometer's case, the supply of battery is a limiting factor, once that was designated just a specific amount of battery units to the caring of this work, and each experiment, that provides just one ASSD observation requires a significant amount of time of performance, and hence, more power consumption. Thereby, since that the increase of the dataset's size is not a viable option, the proper approaches and considerations must be applied into the training and validation phase.

Another crucial aspect observed in the standard dataset, is that the train set is composed of just a few high valued ASSD responses, pictured as purple points, in contrast with the numerous amount of ASSD values ranging below the 0.100 score. Such a discrepancy is identified as outliers in the dataset (OSBORNE; OVERBAY, 2004), once that there is no smooth transition, from an aspect of the dataset, that explains the magnitude of such values. Osborne e Overbay (2004) also assert that, since outliers comply with the increase of error variance and reduce the power of statistical tests, the presence of outliers in a dataset can lead to deleterious effects in statistical analyses. However, although from a general dataset aspect the huge ASSD values from the standard mode represent outliers, yet those values have a meaningful meaning in the real data context. The purpose of the dataset acquired is to characterize a delimited volume of parameters. Thus, regardless of the general dataset aspect, the outlier values convey a prized knowledge of the inclinometer's behaviour, namely the recognition of regions in the delimited volume that lead to an unstable measurement behaviour. Therefore, the regression models employed to be trained with the datasets here presented must be quite flexible in order to perform considerably well, even predicting outliers.

3.6.2 Evaluation metrics employed

According with the discussion in Subsection 2.4.4, there are several evaluation metrics that can be employed to assess the training and validation of a machine learning model. In this work, however, two metrics are deployed in order to remark different aspects of the regression models' performance.

There are two main characteristics of the data that consist the standard and repeated operation mode's dataset. First, especially for the standard operation mode's dataset, the vectors of dependent variable present data scored far outside the general population, where most of the ASSD values are concentrated below the score 0.100, and a few points scatter above this bound. Another aspect concerns the variability of the ASSD values, conveyed through the DOE analyses, which provides an understanding that, depending of the ASSD level, the variability of some performances are considerable high when contrasting with the variability driven by other settings. This phenomenon is remarked when comparing the variability of the trial number eight and one for the sensing y-axis in table 25, where the standard deviation value of the trial eight is more than 140 times greater than the trial one's. This variability concern is not relevant for the training phase, where the algorithm is built to learn the patterns of the training data, but

a situation shall occur in the validation phase, when the built model performs predictions to match the test set. Since there is variability, the predictions will not always match the real value, because this last would be always changing according to its standard deviation. Thereby, if the model had been well fitted without overfitting, the predicted values will be located at a certain close range of the test set value, and this close range is correlated with the variability of that response. The remarking of these two phenomenons in the dataset leads the comprehension that, the evaluation of the goodness of fitting of a model for the case of study of this work must consider a former knowledge of the dataset's patterns, pondering some tolerance of error for the predictions according with the variability of the true ASSD value, in the test set, targeted to be predicted.

In order to assess the degree of similarity between the predicted output, with the true values, the coefficient of determination (also known as R-square, R^2) is employed in this work in order to contrast the variance of the predicted output values to the variance of the true values, as approached in Subsection 2.4.4. Such a metric enables one to numerically ponder the errors of prediction. The more similar are the predicted values to the true values, less is the variance of the errors of prediction, and hence more equal is the true values' variance to the predicted values' variance, according with the equation 14 reasoned by Allen (1997).

Another metric is also employed in this work. This metric was developed by the author in order to consider the variability of each trial in the assessment. Such a metric is nominated as Radius of Tolerance (RT), and is defined as follows. First, one must assign a tolerance value η. Then, the ML model must perform predictions for some set of independent variables of size N (N is the number of rows), with a respective set of dependent variables, where each dependent variable value is a real-valued number represented as y_i. The predictions, also real-valued numbers, represented as $\hat{y}_i$ each, are then compared with the true value y_i. If the absolute difference of y_i and $\hat{y}_i$ is equal or smaller than the tolerance η specified, a variable χ_i assimilated for that prediction is assigned with the value 1. if $|y_i - \hat{y}_i|$ is greater than η, χ_i is assigned with the value 0. This procedure is repeated for all the predictions. Finally, the Radius of Tolerance is defined as the sum of all the assigned χ variables, divided by the size of the independent variables' set N. The Equation 19 portrays the metric's computation.

$$Radius\ of\ Tolerance = \frac{\sum_{i=1}^{N} \chi_i}{N}, \text{ where } \chi_i = \begin{cases} 1, & \text{if } |y_i - \hat{y}_i| \leq \eta \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

One may note that the range of the Radius of Tolerance metric is constrained between 0 and 1. The more closer is the metric value to 1, the more similar are the predictions regarding the true values. The choice of the parameter's η value regularizes the tolerance of the metric. The idea is that η may ponder how much similar is the prediction to a true value that has a variability, and thereby, it is assimilated as a good prediction a value $\hat{y}_i$ that lies within a range of η from y_i, so that $\hat{y}_i \in [y_i - \eta, y_i + \eta]$.

Both the metrics presented so far, R^2 and Radius of Tolerance, are differently employed into the evaluation of this work. At the tunning of the hyperparameters' phase, both the metrics are employed for the standard and repeated mode's dataset. For the validation phase, where the built model predicts the test set, both the metrics are employed for the standard mode's dataset, but just the Radius of Tolerace is employed for the repeated mode's dataset. In this work, the Radius of tolerance is applicable just for true ASSD values below the score of 0.100, with $\eta = 0.010$. The R^2 metric is applicable for all the ASSD values. Such a configuration of metric's deployment is justified due to the general notion, acquired through the DOE approach, that ASSD responses greater than 0.100 may be under influence of more variability than ASSD values bellow the score of 0.100. The assessment of the ML models of this work is more compatible with the reality of the data's aspects through the evaluation by the Radius of Tolerance metric. Had the standard mode's dataset no ASSD values greater than 0.100, the R^2 metric would not be applicable into the evaluation of the validation of the standard mode's test set. During the tunning of the hyperparameters' phase, the deployment of the R^2 metric is considered, for both the operation modes' dataset, in order to complement the perception of the model's goodness of fitting remarked by the Radius of Tolerance. The final assessment of the test set is also achieved, in this work, through graphical comparison between the true and predicted data.

3.6.3 Machine learning employed models

The choice of the ML model to be employed for the solution of a situation must consider the nature of the training data and the phenomenon to be predicted. The inclinometer's ML characterisation proposed in this work must ponder the data's singularities, namely a small dataset, considerable data variance, and outliers. Such data's features are not easily treated. The nature of the inclinometer's characterisation requires a ML approach classified as predictive (or supervised) regression, since the characterisation aims to predict a real-valued value, namely ASSD. At the branch of predictive regression approaches, several distinct strategies are suggestive to be employed in this work, for instance: Neural networks, regression models, ensemble regression and etc. However, just a few of the wide range of strategies are suitable to manage the data's singularities faced in the inclinometer's dataset.

Aggarwal (2018) contrasts the performance of neural networks with traditional machine learning algorithms. The author asserts that neural networks algorithms are capable of, theoretically, learning any mathematical function, and thus describing any phenomenon, if provided a sufficient amount of significant training data. The author complements that even the learning of simple tasks with neural networks often requires an extraordinary amount of data, and that, since traditional machine learning algorithms have a greater ease of model interpretation (i.e. has some bias), they outperform neural networks when handling with small datasets. Such analysis points out that the deployment of neural networks to characterise the inclinometer's behaviour would not be a reasonable choice, since the inclinometer's dataset consists of just a few observations, and it is not affordable to increase the dataset's size. Nonetheless, the analysis

of Aggarwal (2018) also evinces the fact that traditional machine learning models usually trend to achieve a considerable good performance handling with small datasets due to former intrinsic assumptions supposed by the model about the dataset. Such intrinsic assumptions, if too harshly built in the training of the model, may not be advantageous to describe a phenomenon that owns a natural variable response, as it occurs with the inclinometer's dataset, because the algorithm would not be capable of remarking general patterns in the test set for the reason that it would be too bind to the former assumptions consolidated in the training set. In another words, the model would have high bias and high variance (Subsection 2.4.5).

Table 9 – Regression machine learning models briefly experimented in order to observe the performance with the inclinometer's train set.

Model observed	**Performance**
Linear regression	Unsatisfactory
Polynomial regression	Unsatisfactory
Piecewise linear regression	Unsatisfactory
Multivariate adaptive regression splines	Unsatisfactory
Bayesian linear regression	Unsatisfactory
Gaussian process regression	**Satisfactory**
Ridge regression	Unsatisfactory
Support vector machine regression	Reasonable
Nu-Support vector machine regression	Reasonable
LOESS regression	Unsatisfactory
Decision tree regression	Reasonable
Random forest regression	Reasonable
XGBoost regression	**Satisfactory**
K-nearest neighbors regression	Reasonable
Weighted K-nearest neighbors regression	**Satisfactory**
Radius neighbors regression	Reasonable

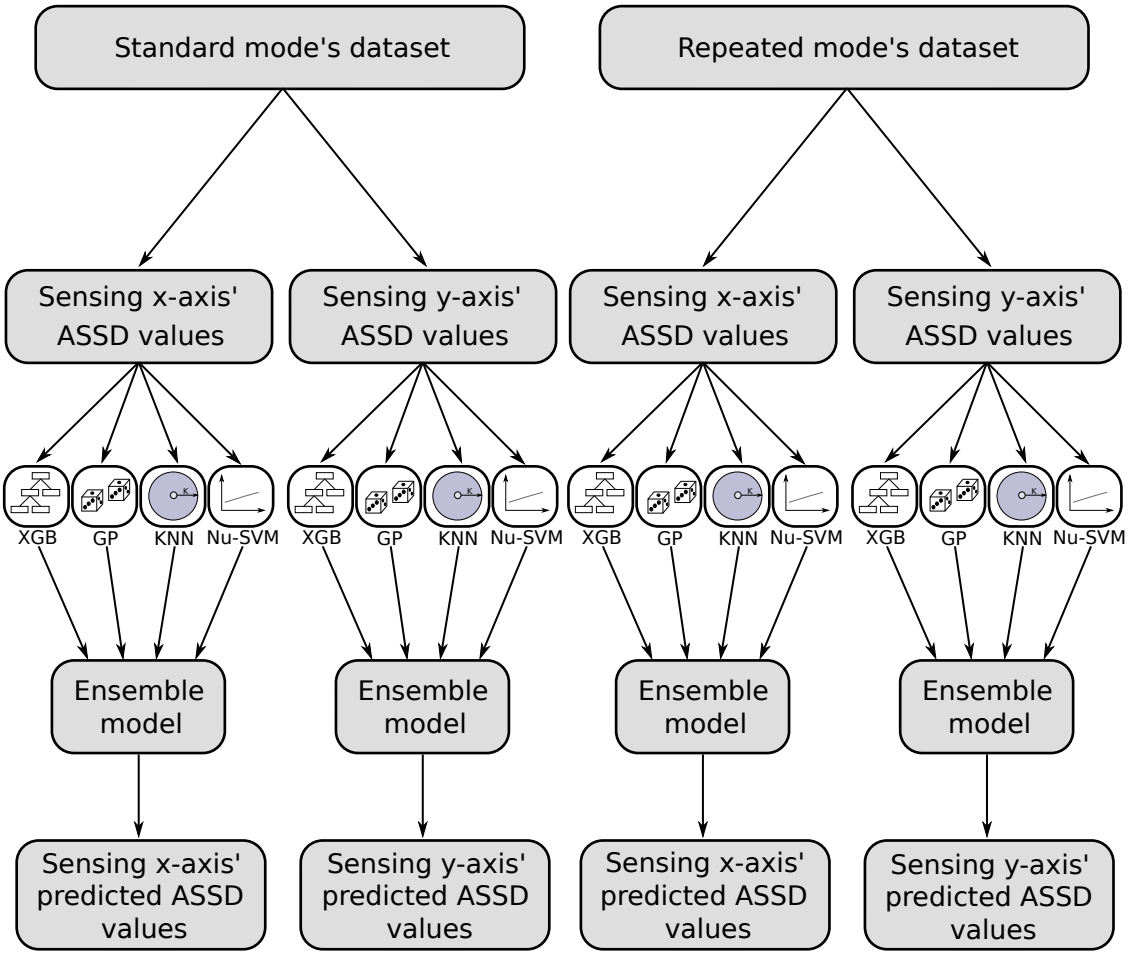
In order to exploit the desirable feature of traditional regression machine learning models (also known as base learners) of to be capable of yielding good performances with small datasets, and yet do not endure the drawbacks of the model's former data assumptions of the inclinometer's dataset, the strategy of ensemble learning is found as a viable solution. In machine learning topics, ensemble learning consists of a collection of algorithms that tries to construct a set of learning models and combine them, aiming to enhance the predictive performance (Subsectiction 2.4.10). In this work, was employed a heterogeneous combination learning technique of the type *Ensemble averaging*. Such approach was explored in this work to train, with the same train set, four base learners whose the learning algorithms are as different as possible among themselves. The procedure of choosing the base learners was carried out by empirically observing the general

performance of several traditional ML models to predict the inclinometer's dataset. The list of models observed was elaborated considering the experiences reported by Kamath e Fan (2018) and by the datascience community, concerning the training of ML models with small datasets. The Table 9 lists the regression base learners briefly experimented and observed for the inclinometer's characterisation prediction performance. The trying of the models was carried out through simply training the algorithm, without any other enhancing technique, and predicting the very train set that was employed to train it.

Due to the singularities of the iclinometer's dataset, just a few among the models experimented had shown satisfactory performance, namely Gaussian process regression (GP), XGBoost regression (XGB) and Weighted K-nearest neighbours regression (WKNN). The satisfaction of the models was assessed through the R^2 metric. According to Zhou (Zhou), it is a general belief that a favourable ensemble model must be composed by base learners as accurate and diverse as possible. Thereby, since the three algorithms that had shown satisfactory performance work through distinctly learning principles regarding themselves, they were chosen to compose the ensemble regression strategy. The XGB regression is substantiated as a decision tree based learner, GP regression learns patterns to perform predictions from a probabilistic approach, and WKNN regression is a nonparametric based algorithm (Subsection 2.4.9). The ensemble learning approach deployed in this work consists of training each one of the chosen models with the same train set, and deploy the built models to predict the same set of vector of independent variable, namely the test or validation set. For each vector of independent variable, the ensemble prediction is the average of the predictions performed by each algorithm for that vector of independent variable, as depicts the Equation 17. In order to provide more variability into the ensemble model so far presented, it was added a fourth base learner to compose the ensemble, namely Nu-Support vector machine regression (Nu-SVM), which is a variant of the classical SVM. Such model, although had had not a satisfactory, but rather reasonable performance, may improve the ensemble prediction accuracy for unobserved datasets, once that the Nu-SVM is a linear based algorithm and thereby might provide significance to the perception of linear aspects of the inclinometers' dataset, diminishing, therefore, the variance and bias of the ensemble.

Since each operation mode of the inclinometer has its own singularities, as for instance a particular number of configurable parameters and distinct interactions among parameters (as remarked by DOE approach), and each mode of operation leads to responses regarding two sensing axis whose the performances have been proved to be distinct between themselves, the machine learning approach consists of assigning for each sensing axis, belonged to each operation mode, an ensemble learning model, in order to characterize the responses for that sensing axis, as schematized in Figure 25. Thereby, the ensemble learner concerning each sensing axis must have its own built base learners with the hyperparameters enhanced for that specific sensing axis, so that the machine learning approach is consisted of 16 built base learners, grouped into four ensemble learners, to perform predictions. Wherefore, since the ensemble strategy employed

Figure 25 – Employment of the ensemble learning strategy into the inclinometer's characterization. Each sensing axis' responses, achieved through the setting of each operation mode, is used to train a unique ensemble model, which is composed of four different base learners whose the learning algorithms are distinct among themselves.



Source: Author.

deploys four regression models, the final ensemble prediction of one vector of independent variable is the average of the base learners' prediction, as portrays the Equation 20.

$$\hat{y} = \frac{\hat{y}_{XGB} + \hat{y}_{GP} + \hat{y}_{WKNN} + \hat{y}_{NuSVM}}{4} \tag{20}$$

To summarize, due to the inclinometer dataset's singularities, namely small amount of data, outliers and considerable data variance, traditional machine learning algorithms, composed of just one base learner, and neural networks, are not a viable option to constitute the ML characterization approach of the inclinometer. In order to consider the data's particularities mentioned, it is employed the strategy of heterogeneous combination learning technique of the type ensemble averaging. Such method may achieve better performance since it is designed to be more accurate, have lower bias and lower variance than single base learners (ZHOU, 2012).

The idea is that, since each base learner owns a different learning algorithm from the others, and thus each of them has a unique bias, the final ensemble prediction might consider the different mathematical and logical aspects of the complex inclinometer's dataset under analyses and thereby achieve better predictions. The final structure of the inclinometer's ML approach is presented in Figure 25, where each sensing axis belonged to each operation mode has its own ensemble model, consisted of four built base learners. All the base learners employed in this work are provided by the open source library developed by Pedregosa et al. (2011) in python 3 programming language.

3.6.4 Training of the models

The machine learning approach consists of the characterization of each sensing axis' responses, for each operation mode, through the training of individual ensemble learning models, composed of four base learners, in order to predict the ASSD value for any combination of parameter at any mode of operation, as describe Figure 25. Since a unique ensemble model is assigned to characterize the performance of one sensing axis at one specific operation mode, each ensemble model must be trained and have its hyperparameters settled individually. Consequentially, a base learner will have hyperparameters settled similarly, but still different, to its respective base learner belonged to other ensembles. Thereby, considering that the training phase procedure deployed to train one base learner is exactly the same deployed to train the others, the following explanations describe the training, tunning and evaluation processes of one base learner.

A usual procedure to train and evaluate the training of a machine learning model consists of the dataset fragmentation into training, validation and test set (GOODFELLOW et al., 2016). So far, the datasets presented in Subsection 3.6.1 regard just the train and test set. No validation set was presented. Since the inclinometer's datasets are consisted of a small amount of data, to slice the train or test set in order to compose a third set of data to consist the validation set is not affordable. No portion of the test set shall be deployed into the training phase, since the very purpose of the test set is to provide unseen observations to attest the ML model. The train set also can not be sectioned, because the training step could be impaired due to lack of enough data. Therefore, in order to evaluate the training performance referent to one set of hyperparameters, the strategy of Leave-One-Out (LOO) was employed to leverage a validation step.

The LOO is a strategy to evaluate the training performance of a ML model for a given hyperparameter set. The technique consists of training the model with the entire dataset, save for one data point (i.e. one independent and dependent variable vector), then deploying the built model to predict the held data point and evaluate this single prediction through the chosen metrics. This procedure is applied for the entire dataset repeatedly, so that at each iteration one different data point is held, the model is trained with the remnant data and the held point is predicted and evaluated, until every dataset's point has been held once. The Algorithm 1 present the LOO procedure deployed in this work, evaluated by the metrics R^2 and Radius of Tolerance

Algorithm 1: Leave-One-Out algorithm deployed in the training of a base learner.

```
dataset_complete = read_dataset( "dataset.csv" );
size_dataset = get_size( dataset_complete );

list_of_R2_trainset = create_empty_list();
list_of_R2_validationset = create_empty_list();
list_of_predicted = create_empty_list();

tol_eta = 0.010;
i = 0;

while i < size_datset do
    train_set = dataset_complete.drop_line_at(i);
    validation_set = dataset_complete.at(i);

    model = fit_model( train_set );

    train_predicted = model.predict( train_set );
    validation_predicted = model.predict( validation_set );

    r2_train = r2_score( train_set, train_predicted );
    r2_validation = r2_score( validation_set, validation_predicted );

    list_of_R2_trainset.append( r2_train );
    list_of_R2_validationset.append( r2_validation );
    list_of_predicted.append( validation_predicted );

    i = i + 1;
end

Radius_of_Tolerance = RT_score( dataset_complete, list_of_predicted, tol_eta);
R2_train = average( list_of_R2_trainset );
R2_validation = average( list_of_R2_validationset );
```

(RT in the algorithm). One may note that the evaluation of one held data point is stored in lists, and at the end of the LOO performance all the dataset's points will have been evaluated and the lists containing the result of each point's evaluation will be fulfilled. The final assessment of the LOO performance for one given hyperparameter set is achieved through the average of the R^2 scores concerning the train set predicted and the validation set (i.e the held points) predicted. For each LOO's iteration, the held point predicted is stored in a list, in order to, at the end of the algorithm, the Radius of Tolerance metric can be applied to assess the LOO performance according to the given η tolerance specified.

Since the LOO strategy is applied to assess the model's performance of just one hyperparameter set, the procedure of finding the best hyperparaemter's settings for just one ML

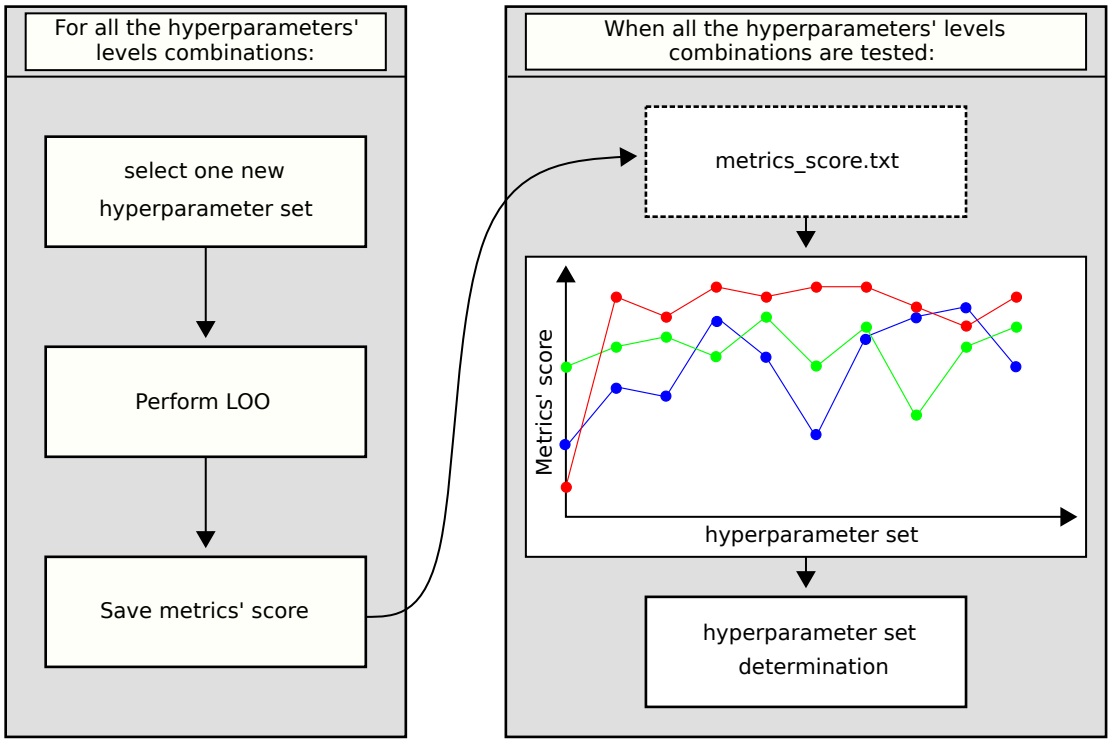
model requires a considerable time. The LOO approach is an efficient methodology to leverage a validation step, since it ponders the prediction of the entire dataset, as like the validation set were the dataset itself, and due to this applicability the LOO is advantageous to be implemented in the evaluation of small datasets, as did Kamath e Fan (2018).

An important aspect that must be considered at the training phase, is whether the ML model to be built must or not be trained with the dataset scaled into some scaling metric (subsection 2.4.3). The specification of the model's necessity of feature scaling is usually specified in the model's API. For the base learners employed in this work, the GP, WKNN and Nu-SVM require feature scaling. For these algorithms, the train set is always scaled with the Standardization score, and the model is trained with the scaled set of data. Consequently, after the model is built, the independent variables to be predicted must be scaled with the same scaling score built for the independent variables deployed into the training step, so that the algorithm may interpret the independent variables to be predicted with the same scale whose the model was built. Therefore, the predicted dependent variable is also a scaled result, according with the scaling score built for the dependent variables of the training step, and in order to one be able to interpret the predicted results in the original scale, they must be reversed scaled by the dependent variable's scaling score built at the training step.

In order to train a machine learning model and reach an optimum performance, the algorithm's hyperparameters must be so settled, that the model is capable of efficiently remarking the patterns of the training data (subsection 2.4.8). Aiming to achieve the best performances with the ensemble learner, the hyperparameters of each base learner must be so settled that each base learner stands for the best performance at the validation set. Since in this work the validation set is leveraged by the LOO strategy, for each ensemble model, each base learner must have had its performance driven by the LOO algorithm for each hyperparameter set tested.

Each base learner has its own specifics hyperparameters. For each hyperparameter, dozens of values were chosen to be experimented, so that the final amount of combinations of hyperparameters evaluated for each base learner could reach hundreds of thousands. In this work, the hyperparameter's levels to be assessed, concerning one base learner, were chosen according to the recommendations and specifications of the model's performance described in the model's API. Once that the hyperparameters' levels to be assessed were chosen, the procedure to experiment all of the hundreds of thousands of possible combinations were structured as a full factorial design, where each model's hyperparameter is a factor, and the evaluation of each performance driven by one set of configuration of hyperparameters was carried out by the LOO approach, as depicted in Figure 26. This process of optimization of the hyperparameters was conducted automatically by written python scripts, so that each hyperparameter set and the respective metrics' scores achieved through LOO were stored and saved in text files. Concluded the hyperparameters' optimization process, the metrics' scores stored were plotted, and the hyperparameter set, whose the performance achieved the higher values at the metrics Radius of Tolerance and R^2 of the validation set predicted, was chosen as the optimum hyperparameter set

Figure 26 – Hyperpatameters' optimization procedure. At the beginning of the process, all the hyperparameters' levels combinations are experimented and their metrics' scores stored in a text file. Afterwards, the scores are plotted and evaluated in order to determine the final hyperparameter set that will be adopted by a base learner.



Source: Author.

for that base learner belonged to one of the four ensemble learners.

The chosen hyperparameter set adopted to be the hyperparameters' levels values of a base learner was determined first pondering performances that achieved higher scores at the Radius of Tolerance metric. Subsequently, the R^2 scores for the validation set predicted were evaluated among those with a high Radius of Tolerance score. Afterwards, R^2 scores concerning the train set predicted were evaluated among those performances with a high score at Radius of Tolerance and R^2 of the validation set predicted. For this last score, differently from the first two, were considered as recommended hyperparameter sets to be adopted, those whose R^2 score of train set predicted were more similar with the R^2 score of the validation set predicted. Such consideration has been done in order to avoid overfitting, since if the score achieved at the validation set were notably lower than the score achieved at the train set, it means that the model was too bounded to the train set's patterns and has a lower generalization to accurate predict unseen observations (Subsection 2.4.6 and 2.4.8). Considering these criteria, the hyperparameter sets of a single base learner were chosen.

The procedure of training and determine the hyperparameters' values so far presented concerns just for the building of one single base learner. Such procedure was deployed in the

training phase of all the 16 base learners employed in this work. Once that the base learners that consist of an ensemble learner are trained and optimized, the ensemble model can be attested with the LOO algorithm in order to remark the ensemble's performance at the validation set. If the metrics' scores achieved by the ensemble were satisfactory for the validation set's prediction, the ensemble is ready to be finally assessed predicting the test set. If the scores achieved were not satisfactory, then the hyperparameters of the ensemble were again pondered and reevaluated.

The strategies employed in this work, such as ensemble learning and LOO, are recommended approaches to handle with the particularities of a small dataset. Even though, the time required to train and evaluated all the base learners through the LOO strategy combined with the hyperparameters' optimization process was considerable. All the training and testing performances were carried out through the author's personal computer. Were the datasets' size greater, a requirement of a machine with more computation power would be necessary, in order to speed up the machine learning computations. The scores achieved by each base learner in the validation set, set with the optimum hyperparameters set, are presented in the Section 4.1.

3.6.5 Validation of the models

The validation of the ensemble learning models built to predict the ASSD's performance of the inclinometer device for any combination of parameters inside a delimited volume of parameters, given the operation mode, was carried out through the performance achieved at the prediction of the model's respective test set. The prediction performance is evaluated through the scores achieved at the Radius of Tolerance metric with tolerance $\eta = 0.010$ and, just for the standard operation mode, R^2 score of the test set predicted. The performances achieved in the test set's prediction are presented in section 4.2.

4 RESULTS AND DISCUSSION

In this chapter is presented the results of the machine learning approach relied on the DOE technique, reasoned to characterize the response of the inclinometer device, developed by the Large-Scale Metrology team, for all the measurement parameters' combination in a delimited volume of parameters. First, is approached the training scores achieved by the ensemble regression models assigned for each sensing axis' responses, for two operation modes: Standard and repeated. Following, the trained models are undergone to a validation phase, where each built ensemble model predicts the test set, and these predictions are evaluated in order to attest the accuracy and reliability of the built ensemble models, and hence, the ML approach. Afterwards, the built algorithms are employed to fully characterize a volume of parameters, and the performances predicted are contemplated and discussed. To conclude, the entire methodology is assessed and discussed, pondering the effectiveness, advantages and drawbacks of such technique developed to complement the DOE characterization.

4.1 MACHINE LEARNING MODEL'S TRAINING SCORES

In the Section 3.6 was presented the algorithms and approaches whose the machine learning characterization substantiate itself. In this section, the performance of the ensemble models' training phase, approached in Subsection 3.6.4, is assessed through the evaluation metrics reasoned in Subsection 3.6.2.

Table 10 – Scores achieved in the training phase of the base learners and the ensemble regression, concerning the standard operation mode. The table present the scores achieved by the models trained for the sensing x-axis' characterization, in the left side, and for he sensing y-axis' characterization in the right side.

Standard Operation Mode	**Model**	**RT**	**R^2 Val**	**R^2 Train**		**Model**	**RT**	**R^2 Val**	**R^2 Train**
Sensing x-axis	XGB	0.63	0.95	0.99	**Sensing y-axis**	XGB	0.75	0.90	0.93
	GP	0.71	0.77	0.97		GP	0.56	0.83	0.99
	WKNN	0.50	0.62	0.80		WKNN	0.50	0.59	0.81
	Nu-SVM	0.65	0.35	0.42		Nu-SVM	0.73	0.35	0.45
	Ensemble	0.63	0.82	0.79		Ensemble	0.68	0.86	0.80

The Table 10 presents the scores achieved by two built ensemble models, one for the characterization of each sensing axis, concerning the standard operation mode. The left side

of Table 10 stands for the scores related to the sensing x-axis' models, whereas the right side stands for the scores regarding the sensing y-axis' models. Each line of scores concerns either a base learner's performance, or the performance of the ensemble whose the base learners consist. The column RT stands for the score achieved at the Radius of Tolerance metric with $\eta = 0.010$, comparing the predictions of the validation set (i.e. the LOO predictions) with their true values. The column $R^2\ Val$ regards the scores achieved in the R-Square metric for the very predictions of the validation set deployed in the evaluation of the RT metric, comparing the predicted values with their true values. Finally, the column $R^2\ Train$ stands for the score achieved in the R-Square metric by the predictions of the very train set employed to train the model that predicts it, comparing the predicted training values with their true values. The RT metric was applied to assess the predicted ASSD values whose the true values are found below the rank of 0.1, complying with the conclusions led by the DOE approach, reasoned in Subsection 3.6.2. The $R^2\ Val$ and $R^2\ Train$ assess the validation and train set, respectively, as enlightened in Subsection 3.6.2. The scores presented in Table 10 were achieved by base learners set with an optimum hyperparameter set, determined by the procedure described in Subsection 3.6.4.

Assessing the performance of the base learners of Table 10, one may observe that some models could better remark a specific type of patterns than others. For instance, the algorithm GP, for the sensing x-axis, could remark quite distinctly the patterns for ASSD values below the rank of 0.1, however its performance at the $R^2\ Val$ is overtaken by the XGB algorithm. As matter of fact, the fraction amount of ASSD values in the sensing x-axis' validation set that lay below the rank of 0.1 is $38/64$. Still in the sensing x-axis, the RT score achieved by the Nu-SVM algorithm conveys that it could considerably well remark lower ASSD values' patterns, however it had no great performance to observe the patterns of higher ASSD values. The WKNN model achieved a reasonable good performance at both the RT and $R^2\ Val$ metrics. Focusing in the $R^2\ Train$ scores achieved by the base learners for the sensing x-axis, one may observe that, save for the GP algorithm, the remnant models could achieve a $R^2\ Train$ value quite similar with the $R^2\ Val$. The fact that the GP algorithm achieved a quite different value at the $R^2\ Train$ score from the $R^2\ Val$, may indicate that the model is overfitted (Subsection 2.4.5, 2.4.6 and 2.4.8). However, when combining the performance of each base learner, stood for the sensing x-axis, into an ensemble, the ensemble's performance ponders the perspectives of each base learner and provides a general response that is quite satisfactory for the three evaluating metrics here employed. The ensemble performance of the sensing x-axis of Table 10 does not achieve an outstanding performance in the $R^2\ Val$ score, like the XGB, neither a distinctly good accomplishment in the RT metric, as did the GP algorithm, but yet, the ensemble could balance the bias of each model, achieving a good perspective of the train and validation set's patterns. One may also observe, still for the sensing x-axis performances, that the ensemble's $R^2\ Val$ score could even outweigh the $R^2\ Train$ score, such fact that strongly suggests that the ensemble is not overfitted.

Assessing the performance of the base learners from Table 10 for the sensing y-axis'

characterization, one may observe that the scores achieved by the models referent to the sensing y-axis are quite similar to those that stand for the sensing x-axis, but with some slight differences. As matter of fact, the fraction amount of ASSD values of this sensing axis' validation set that lay below the rank of 0.1 is $48/64$. The XGB model could achieve outstanding scores for both the RT and $R^2\ Val$ metrics. The GP base learner achieved a considerable good score in the $R^2\ Val$ metric, but reasonable good in the RT metric. The WKNN achieved a reasonable performance for both the metrics. As it occurred for the sensing x-axis' Nu-SVM base learner, also for the sensing y-axis predictions the model could notably well remark the pattern of lower ASSD values, but had not an impressive score for the $R^2\ Val$ metric. The scores achieved in the $R^2\ Train$ were, again save for the GP model, quite similar to the scores achieved in the $R^2\ Val$ metric. As pointed for the sensing x-axis scores' assessment, the fact that the GP model achieved a particular higher score in the $R^2\ Train$ metric than in the $R^2\ Val$, may suggest that the model is overfitted (Section 2.4.5, Subsection 2.4.6 and 2.4.8). Yet, the combination of the base learners into an ensemble regression could leverage considerably good scores for the three metrics. Again, the ensemble regression could even achieve a better score in the $R^2\ Val$ metric than in the $R^2\ Train$ metric, such fact that may strongly suggest that the model is not overfitted.

The two ensemble models of the standard operation mode's characterization achieved similar scores, although the model for the sensing y-axis' characterization had presented performances slightly better. The performances at the RT metric were not outstanding, but yet were considerable reasonable. The performances at the $R^2\ Val$ metric were considered satisfactory, once that this metric assesses the quality of prediction of the entire validation, and the present dataset is consisted of several particularities as outliers and small amount of training data. Hence, a R^2 score above 0.80 at the validation set suggests that the ensemble could well learn most of the dataset's patterns. From such analyses, one may assume that the ensemble regressions for the standard operation mode's characterization had a good intuition of the patterns that rule ASSD values below the rank of 0.1, and have pretty well learned the patterns of the dataset/validation set seen from a general aspect.

The Table 11 is organized exactly as the Table 10 , but concerns for the training scores achieved by the base learners and ensemble models charged to characterize the repeated operation mode. The RT, $R^2\ Val$ and $R^2\ Train$ metrics were here employed as it was for the standard operation mode models' metrics, save for the applicability of the RT metric, which for the analysis concerning the standard operation mode is just applicable for ASSD values bellow the rank of 0.1, and here, for the repeated operation mode, is applicable for the entire dataset/validation set, since the repeated operation mode's dataset is fully consisted of ASSD values below or equal to the rank of 0.1.

Assessing the base learners and the ensemble regarding the sensing x-axis' characterization, left side of Table 11 , one may observe that all the base learners could achieve satisfactory scores in the three metrics. The performance of the XGB and GP algorithms were outstanding for the remark of the dataset/validation set's patterns, and the WKNN and Nu-SVM

Table 11 – Scores achieved in the training phase of the base learners and the ensemble regression, concerning the repeated operation mode. The table present the scores achieved by the models trained for the sensing x-axis' characterization, in the left side, and for he sensing y-axis' characterization in the right side.

Repeated Operation Mode	**Model**	**RT**	**R^2 Val**	**R^2 Train**		**Model**	**RT**	**R^2 Val**	**R^2 Train**
Sensing x-axis	XGB	1.00	0.97	0.98	**Sensing y-axis**	XGB	1.00	0.91	0.97
	GP	1.00	0.96	0.99		GP	1.00	0.89	0.99
	WKNN	0.93	0.89	0.93		WKNN	0.95	0.73	0.85
	Nu-SVM	0.91	0.90	0.95		Nu-SVM	0.91	0.67	0.83
	Ensemble	1.0	0.95	0.96		Ensemble	0.97	0.85	0.91

could pretty well remark it. The ensemble of the base learners achieved, as well the base learners, satisfactory scores. The scores achieved at the $R^2\ Train$ metric, by the base learners and the ensemble, conveys the suggestion that none of the algorithms was overfitted, since the score at the $R^2\ Train$ metric was pretty similar to the score achieved at the $R^2\ Val$ metric.

Similarly were the scores achieved by the base learners and the ensemble regression concerning the sensing y-axis' characterization, presented on the right side of the Table 11. The XGB and GP algorithms could achieve satisfactory scores at the three metrics. An interesting phenomenon occurs for the scores achieved by the WKNN and Nu-SVM models, where both could pretty well remark the patterns of the dataset/validation set, but the reasonable scores at the $R^2\ Val$ metric, when comparing with the scores achieved at the same metric by the XGB and GP algorithms, conveys the suggestion that the quality of the patterns learned by the WKNN and Nu-SVM was not so accurate as it was for the GP and XGB algorithms. In other words, it is suggested that the absolute difference between the true and predicted ASSD values were more close to the tolerance η defined in the RT metric, for the WKNN and Nu-SVM, than the same absolute difference achieved by the XGB and GP models. Nevertheless, the ensemble of the base learners charged to characterize the sensing y-axis' performance could achieve satisfactory scores at the three evaluation metrics. The scores achieved by the base learners and the ensemble at the $R^2\ Train$ metric were quite similar to the scores achieved at the $R^2\ Val$, save for the GP model, which achieved a not so negligible difference between the mentioned metrics, which may suggest that the base learner is slightly overfitted.

The ensemble regressions of the repeated operation mode's characterization achieved satisfactory scores at the validation set, although scores achieved by the ensemble concerning the sensing x-axis' characterization were better than the scores achieved by the ensemble which regards the sensing y-axis' characterization. The performances at the RT metric were

considered outstanding, and the performances at the $R^2\ Val$ metric were considered satisfactory. Although the repeated operation mode's dataset contains fewer observations than the standard operation mode's dataset, the scores' achievement was notably better at the repeated mode characterization's training than for the standard mode. Such phenomenon may be justified due to the fact that the standard operation mode's dataset presents a more complex pattern of ASSD performances than the repeated operation mode's dataset. This last is composed of just two independent variables, whereas the first is consisted of three independent variables. Also, the standard operation mode's dataset has a scale of ASSD values pretty diverse, even containing outliers, such that the RT metric could not even be applied to assess the entire set of data, but just those whose the true ASSD values lay below the rank of 0.1. Such particularities are not trivial to bypass, so that even a flexible approach, as it is the ensemble regression, may have complications with the training and fitting of such dataset. The repeated operation mode's dataset is simpler, and each of its ASSD values lay below or equal the rank of 0.1. Hence, the training with such batch of data is more feasible with the ensemble approach.

4.2 MACHINE LEARNING MODEL'S VALIDATION

In Section 4.1 was presented the scores achieved by the trained models, with the best hyperparameters set settled, for the predictions of the validation and train set. The performance at the training phase (Subsection 3.6.4) leverages an assessment of the goodness of the algorithms' fitting. In this Section, it is presented the performance and scores achieved by the ensemble models for the prediction of unseen observations (i.e. data points that the ensemble models have not been trained to achieve good predictions' performance), namely the test set approached in Subsection 3.6.1. The performance of the ensemble models at the test set's prediction attests, in this work, the reliability and accuracy of the machine learning approach to characterize the inclinometer's behaviour for any measurement parameters' combination constrained in a delimited volume of parameters.

The deployment of the ensemble models to predict the test set was carried out as follows. First, the ensembles of base learners were trained with the best hyperparameter set. Then, the built models were deployed to predict all the possible combinations of the parameters' values presented in Table 6, for the standard operation mode's measurement, and Table 8, for the repeated operation mode's measurement. Afterwards, the predictions were evaluated comparing their values with the true values presented in the test set. The evaluation was accomplished following the metrics' deployment reasoning developed in Subsection 3.6.2.

The Table 12 conveys the scores achieved by each ensemble model, which stands for one sensing axis and one operation mode, concerning the performance at the predictions of the test set. The tolerance value η of the Radius of Tolerance metric, deployed in the test set's assessment, was 0.010 for both the operation modes. For the predictions concerning the standard operation mode, the RT metric was applicable just for true ASSD values below the rank of 0.1,

Table 12 – Scores achieved by the ensemble models, which consist the machine learning characterization approach, at the performance of the test set's prediction.

	Standard		**Repeated**
	RT	$\mathbf{R^2}$	**RT**
Sensing x-axis	0.42	0.70	0.90
Sensing y-axis	0.71	0.83	0.96

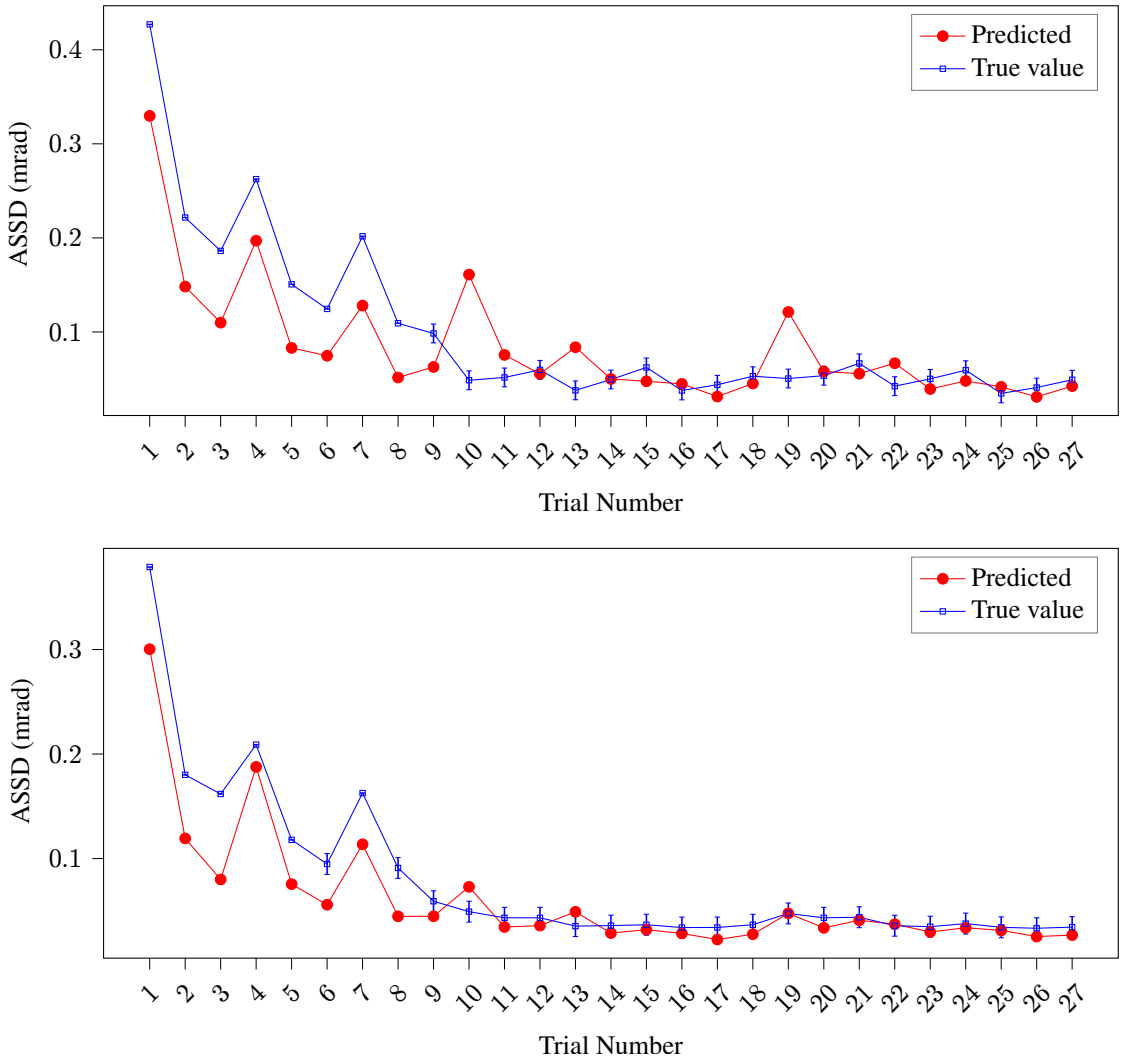
as reasoned in Subsection 3.6.2. The fraction amount of predictions evaluated with the RT metric for the sensing x-axis, at the standard operation mode, was $19/27$, whereas the fraction amount for the sensing y-axis, at the same operation mode, was $21/27$. The R^2 metric was applicable for all the test set's predictions concerning the standard operation mode, and the RT metric was applicable for all the test set's predictions concerning the repeated operation mode.

For the test set's predictions concerning the standard operation mode, one may conclude, through the observation of the Table 12, that the ensemble model approach could considerably well remark the general patterns of the test set, since the score achieved at the R^2 metric was considered satisfactory, for the ensemble regarding the sensing y-axis, and reasonable good, for the ensemble concerning the sensing x-axis. The models' performance of predictions of ASSD values below the rank of 0.1, assessed through the RT metric, was considered unsatisfactory for the predictions regarding the sensing x-axis, and satisfactory for the predictions concerning the sensing y-axis. One may also note that, still assessing the standard operation mode, the ensemble model that regards the sensing y-axis achieved better scores than the ensemble model that stands for the sensing x-axis.

In order to visually remark the performance of predictions concerning both the sensing axes, and thereby graphically assess the quality of the predictions, the true values of the standard operation mode's test set were plotted along with the respective predictions, as presented in Figure 27. The abscissa axis stands for the standard operation mode test set's trial number, found in the Table 35 in the Appendice C. The first graphic on top of the Figure 27 stands for the sensing x-axis' performance, whereas the second graphic at the bottom stands for the sensing y-axis' performance. Comparing the true and predicted values of the graphic on top, one may observe that the ensemble regression could reasonably predict the shape of the test set, yet such predicted shape is not totally well fitted into the true values' line. It is also possible to observe, still in the graphic on top, that the ML model had the wrong assumption about the prediction's pattern for some trials number, namely number 10, 13 and 19. Comparing the true and predicted values of the graphic at the bottom, one may observe that the shape of the true values could be well remarked by the ensemble regression, and the predicted values are placed considerably close to their respective true values.

By the means of the models' performance evaluation presented in Table 12 and Figure 27, it is considered that the two ensemble regression models, each one built to characterize one

Figure 27 – Performance of the predictions of the ensemble regression models, which have been trained to characterize each sensing axis at the standard operation mode configuration, at the test set. The graphic on top stands for the performance of the ensemble model built to predict the ASSD score reached by a set of parameters performed by the sensing x-axis, whereas the graphic on bottom regards the sensing y-axis. The blue squares are the true values obtained in real experiments. The red dots are the values predicted for the same parameters that have led to the blue squares.



Source: Author.

sensing axis' measurements, at the standard operation mode, could remark the general aspect of the inclinometer's behaviour with satisfactory accuracy, for the sensing y-axis' measurements, and reasonable accuracy, for the sensing x-axis' measurements. The difference in accuracy achieved by the two ensemble models may be justified by the different aspects of data, particular of each sensing axis. The Figure 22, in Subsection 3.6.1, conveys the knowledge that the sensing x-axis intrinsically reaches higher ASSD values than the sensing y-axis, and through the interpretation of the DOE model performed to comprehend the standard operation mode, it became evident that the higher is an ASSD value, the higher is the uncertainty of that value.

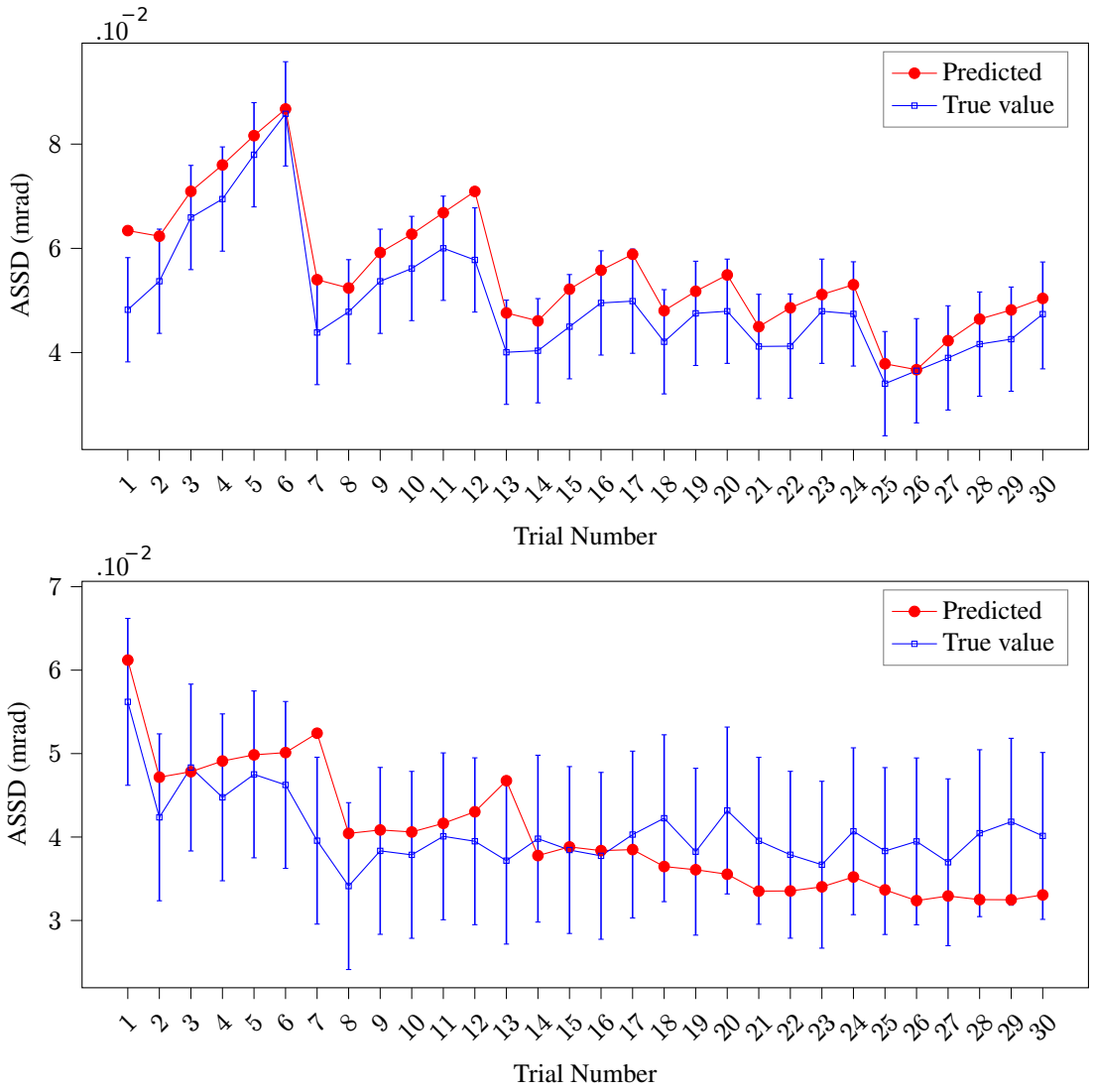
Therefore, it is expected that the machine learning model achieves, at the standard operation mode, an worse accuracy for predictions regarding the sensing x-axis, than for the sensing y-axis', once that one single true sensing x-axis' ASSD value, achieved with a given configuration of parameters, may vary in within a range wider than a sensing y-axis' ASSD value achieved through the same parameters' combination, and hence a sensing x-axis' predicted ASSD value will always have a higher probability to lie more distant of the respective true value than a sensing y-axis' predicted ASSD value, comparing both the sensing axis' measurement performance for the same set of parameters.

The Table 12 also conveys the performance of the ensemble regression models that concern the characterization of the repeated operation mode's measurements at the test set. Since the R^2 metric is not an appropriate approach to evaluate the repeated operation mode's prediction performances, as discussed in Subsection 3.6.2, just the Radius of Tolerance metric is employed in the evaluation of the operation mode's predictions. Thereby, one may conclude by the means of the score achieved by each ensemble, one for the characterization of each sensing axis that stands for the repeated operation mode, that the predictions' performance of the algorithms was considerably accurate.

The Figure 28 provides a visual understanding of how the two ensemble regression models remarked the repeated test set's patterns. The first graphic on top of the figure stands for the sensing x-axis' predicted and test set values, whereas the graphic at the bottom stands for the sensing y-axis'. The abscissa axis of both the graphics regards the trial number of the test set in compliance with the repeated operation mode's test set presented in Table 37, found in the Appendice D. As for the graphics concerning the standard operation mode's predicted and test set values, here the tolerance value $\eta = 0.010$ of the RT metric is evinced with vertical bars at each true value, so that is considered a predicted value acceptable according to RT's η tolerance, one that lies within the range of the respective true value's vertical bar. Assessing the graphic on top of Figure 28, one may observe that the ensemble regression could so accurately remark the patterns of the sensing x-axis ASSD responses, that almost all the predicted values lie within the constraint of the RT's tolerance. In fact, for the sensing x-axis, the ensemble algorithm could satisfactorily remark the shape of the repeated operation mode's test set. Assessing the graphic at the bottom of Figure 28, immediate considerations must be pondered. The sensing y-axis test set's values achieved roughly the same ASSD value. Such phenomenon has already been observed by the DOE assessment of the operation mode in Subsubsection B.3.2, where the ASSD values barely change according to different parameters' combinations around a certain spot of parameters' values. Nevertheless, the ensemble regression model assigned a pattern to the sensing y-axis' performances, and such pattern could lead the predictions to lie within the RT's tolerance, as convey the vertical blue bars that depart from the test set values, although the shape of the predicted values does not well fit to the true values'.

The evaluation of the regression models' performance at the test set is definitive to attest the model's reliability in a real use case. It is relevant to point out that the test set, of

Figure 28 – Performance of the predictions of the ensemble regression models, which have been trained to characterize each sensing axis at the repeated operation mode configuration, at the test set. The graphic on top stands for the performance of the ensemble model built to predict the ASSD score reached by a set of parameters performed by the sensing x-axis, whereas the graphic on bottom regards the sensing y-axis. The blue squares are the true values obtained in real experiments. The red dots are the values predicted for the same parameters that have led to the blue squares.



Source: Author.

both the operation modes, was collected following the same systematic methodology of ASSD values' acquirement, employed to obtain the train set. It is also relevant to evince that the test set was acquired several months later of the train set's acquirement, so that it has no correlation at all with the train set, save for the systematic methodology of experiments, and thereby if the ML approach could achieve good predictions of the test set, the approach would be considered satisfactory. A point that might be pondered is that the exposition and use of the inclinometer device in the time between the acquisition of the train and test set could have modified internal

electronics aspects of the printed circuit board or the MEMS accelerometer ICs, so that subtle features of the device's measurement pattern could have been changed. Such change, if not underlying, is advantageous to attest the capability of generalization of the built ensemble models for their deployment in real use cases.

Concluding, given the particularities of the standard operation mode's dataset, as ASSD values with high variability and outliers, it is considered that the ML model built to characterize the standard operation mode has, pondering in a general aspect, a satisfactory accuracy for the predictions concerning the sensing y-axis, and a reasonable accuracy for the sensing x-axis. It is considered that one may achieve a decent intuition of the patterns of quality of the sensor's measurement, for any configuration of parameters constrained in the train set's volume of parameters. The performance of the built models that concern the repeated operation mode's characterization is considered to have a satisfactory accuracy for the predictions regarding both the sensing axes. The intuition provided by the ensemble models charged to characterize the repeated operation mode is considered efficient for measurements of both the sensing axes.

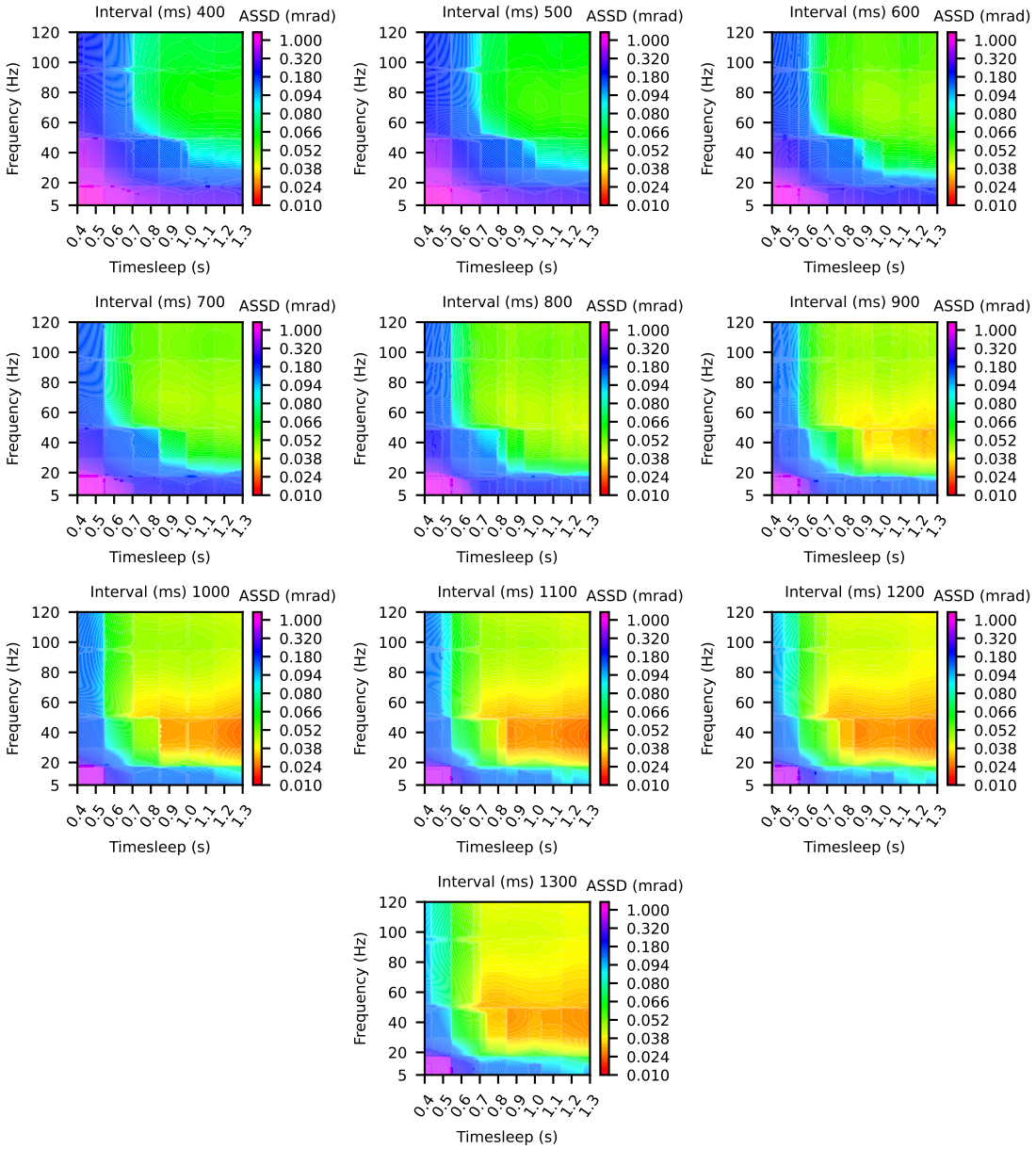
4.3 MACHINE LEARNING INCLINOMETER'S CHARACTERIZATION

In this section is presented the full characterization achievable by the ML approach. Each trained ensemble model is deployed to predict the performances' ASSD values referent to all the possible combinations of parameters constrained into the train set's volume of parameters, according to the model's sensing axis and operation mode that it was trained to predict, found in Table 5 and 7.

The figures 29 and 30 present the full characterization of the quality of measurements performed in the standard operation mode, which is assessed through the ASSD metric, provided by the machine learning approach. The very same two built models attested in the prediction of the standard mode's test set are here employed. The Figure 29 stands for the sensing x-axis' characterization, whereas the Figure 30 stands for the sensing y-axis'. Each coloured scaled plot in the figures stands for the ASSD values' predictions of all the possible combinations of *frequency* and *timesleep* values, when the parameter *interval* is settled to one fix value, where the range of *frequency* varies from value 5 to 120 with a resolution of 1, and the range of *timesleep* varies from value 0.4 to 1.3 with a resolution of 0.01. The colour scale of each plot is the same, and it stands for the ASSD value predicted, where the higher is the number predicted, the worse is the quality of measurement's performance.

Assessing the predictions in Figure 29, one may clearly remark the patterns that the ensemble model had learned, and, afterwards, deployed to perform the full characterization. Focusing on the characterization of the lower intervals, one may observe three main regions of ASSD values. As coloured in faded pink is found the region where the ASSD values reach the highest rank. As coloured in blue is found still a region where the ASSD values are high, but they are gradually decreasing. And as coloured in green is found the region where the

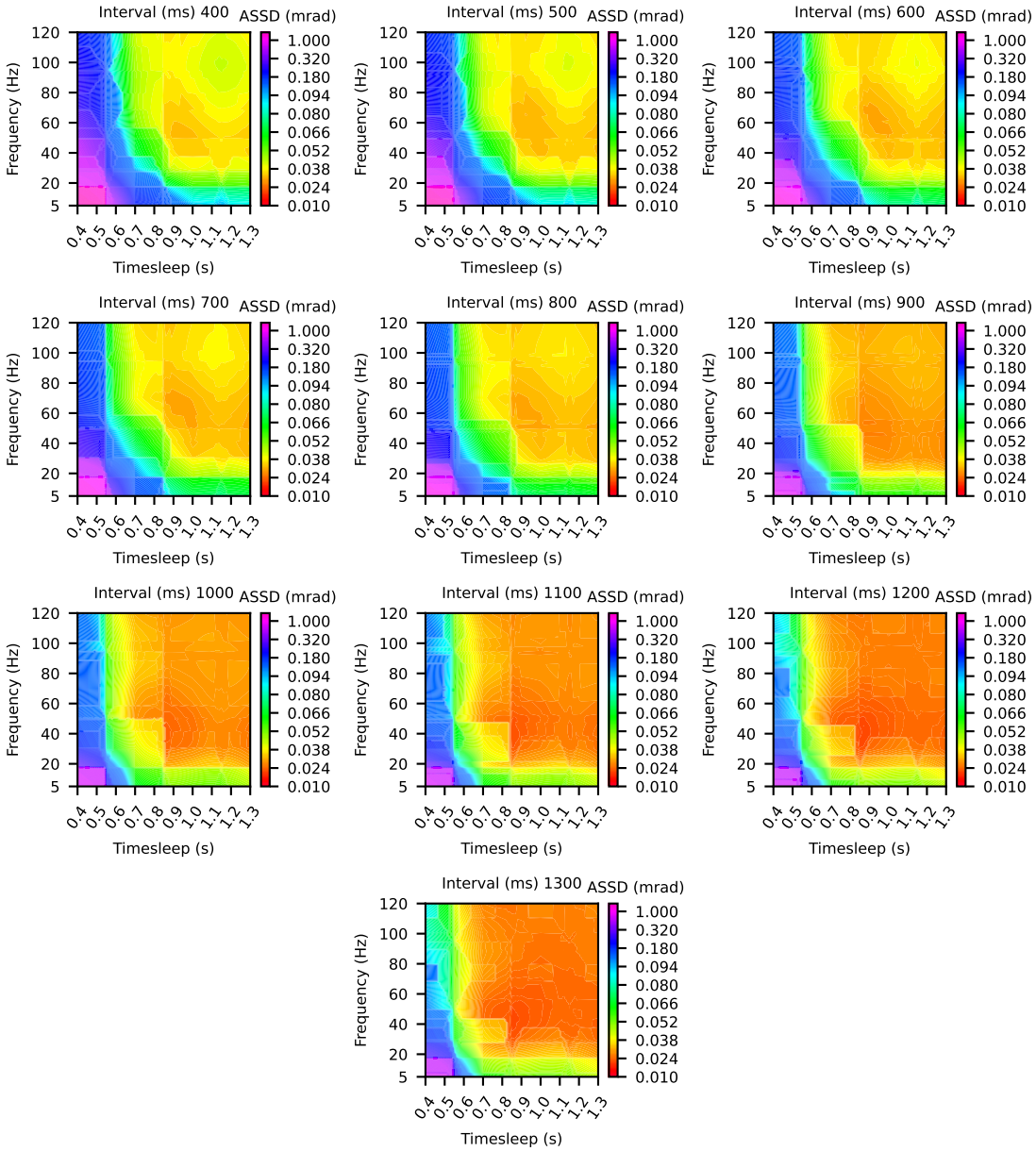
Figure 29 – Full characterization of the standard operation mode for the sensing x-axis' responses.



Source: Author.

measurements are found more stable and with more accurate performance, so that the ASSD values are found below the rank of 0.1. For the lower intervals, the worse performances are achieved when *frequency* and *timesleep* are set to small values, and the performance is improved along with the increasing of these parameters' values. It is interesting to note that, as *interval* increases, the faded pink region becomes smaller, and gradually sets on a small spot of *frequency* and *timesleep* values, namely the smallest values of them, but the algorithm does not extinguish the region for any value of interval. Also, as *interval* increases, the green region expands, and a new region, which is coloured as yellow/orange and stands for the smallest ASSD values, is

Figure 30 – Full characterization of the standard operation mode for the sensing y-axis' responses.



Source: Author.

arisen roughly between the *frequency* range of $[20, 70]$, and *timesleep* range of $[0.8, 1.3]$. Such region reaches the largest size and smallest ASSD values between the *interval* of values 1100 and 1200, so that the region stands for a sweet spot of better performances of the sensing x-axis' measurements. At the *interval* 1300, this region seems to become fading.

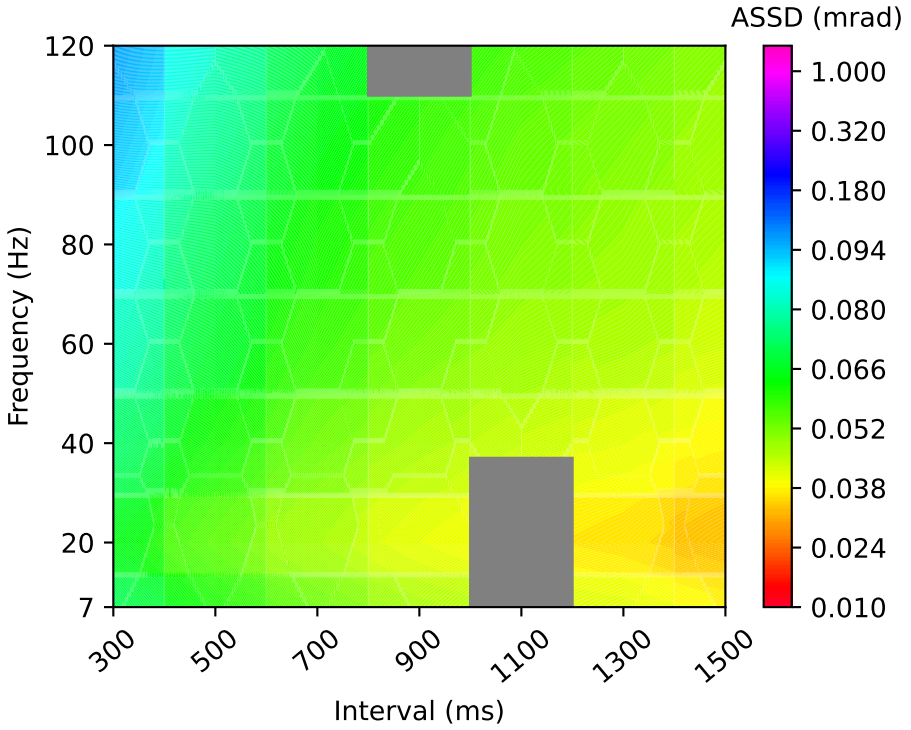
Assessing the predictions in Figure 30, the patterns of the sensing y-axis' performances are evinced. Focusing on the lower intervals, differently of the three regions predicted for the lower intervals in Figure 29, four regions are here found. As in Figure 29, there are a faded pink, a blue and a green region, which stand for the same scale of ASSD values remarked in the

analysis of Figure 29, but a fourth region is arisen, namely a yellow/faint orange coloured region. The contrast with the predictions of the sensing x-axis responses is remarkable. The faded pink region is still concentrated when *frequency* and *timesleep* are set to low values, but such region is smaller than it is in the lower intervals of Figure 29. The blue region is here, in the lower intervals of Figure 30, reduced in size. An interesting comparison with the blue regions in the lower levels of Figure 29 is that here, the blue region does not reach the spot where *frequency* is set to lower values and *timesleep* is set to high levels, but instead of blue, the spot is occupied by a green/faint blue region. Such pattern leads to the conclusion that it is possible to reach ASSD values below the rank of 0.070, in the standard operation mode, for low values of *frequency* and *interval*. Such an accomplishment that is not achievable in the sensing x-axis' measurements. It is also interesting to note that, still assessing the lower intervals' values, the yellow/faint orange region is becoming with a colour of more deep orange as *interval* is increased, and the location of such spot roughly complies with the location of the orange spot observed in the high values of *interval* in Figure 29. The increasing of *interval* leads to a general better performance of measurement, so that the ASSD values are gradually decreasing in a general aspect. The yellow/faint orange region expands in high values of *interval*, so that it almost becomes just with orange colour. In this region, a notable deep orange spot is arisen roughly when *frequency* ranges within $[20, 60]$, *timesleep* ranges within $[0.7, 1.0]$, and *interval* is found between $[1000, 1300]$, so that the region represents a sweet spot of parameters' combinations that leads to an optimum performance of the sensing y-axis' measurements.

The full characterisation achieved by the two ensemble models charged to characterize the standard operation mode could remark different regions, which stand for combinations of parameters spot, that lead to different aspects of the operation mode's performances. One may deploy the figures 29 and 30 in order to choose the parameters' values that lead to the most suitable performance for any application. For instance, if one is interested in saving power consumption, but still wishes to keep the better accuracy as possible in the measurements of both the sensing axis, the most suitable option would be to set *timesleep* to 1.3, *frequency* to 20 and *interval* to 1300, because this configuration would ensure a slow rate of measurement sampling, which would decrease the power consumption, and the performance would still remain with ASSD values below the score of 0.1 in both the sensing axes' measurements. But if the application requires a high rate of measurement sampling, still with the best accuracy as possible in both the sensing axes' responses, an interesting recommendation would be to set *interval* to 400, *frequency* to 120 and *timesleep* to 0.7. Such configuration would ensure short intervals of measurement, with a reasonable good accuracy, but the power consumption would increase, once that the amount of tilt measurement samples, necessary to provide one single final measurement sample (Subsection 3.1.1), would increase. It is also interesting to emphasize that in real performances, the real ASSD values achieved by the inclinometer device would much probably not reach, for a given set of combinations of parameters, the exact value suggested by the characterization, since the ensemble models could well remark the general aspect of the

standard operation mode's patterns of ASSD responses, but the predicted values have mostly not been in compliance with the tolerance of $\eta = 0.010$ of the Radius of Tolerance metric.

Figure 31 – Full characterization of the repeated operation mode for the sensing x-axis' responses.

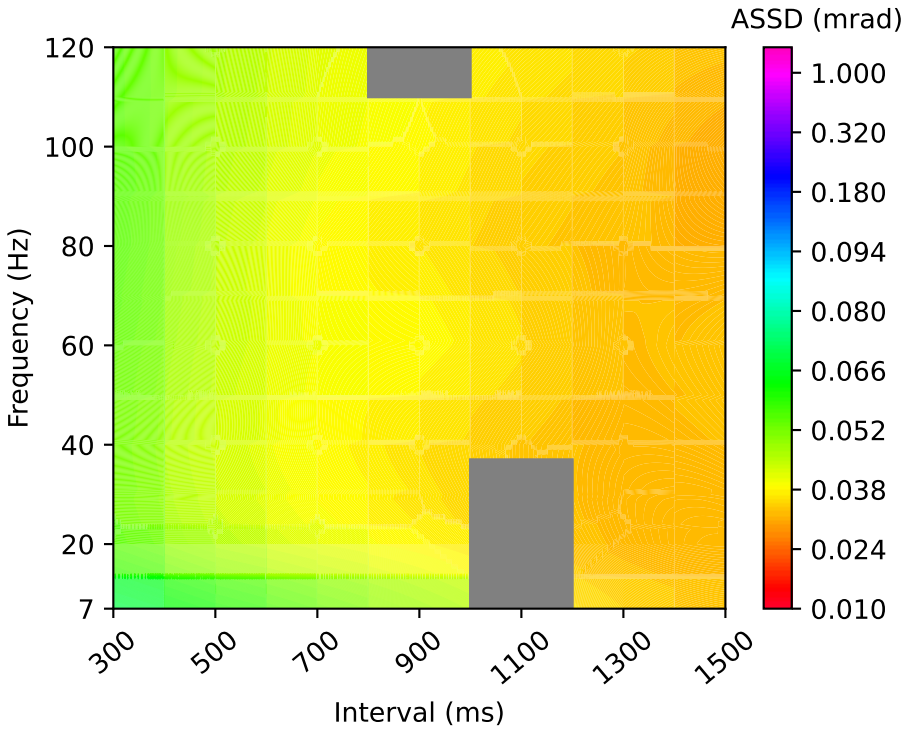


Source: Author.

The figures 31 and 32 present the characterization of the quality of measurements performed at the repeated operation mode, assessed through the ASSD metric, provided by the machine learning approach. Figure 31 stands for the characterization of the quality of measurements performed by the sensing x-axis, whereas Figure 32 stands for the sensing y-axis' quality of measurements. The repeated operation mode's characterization consists of the ASSD predictions of all the possible combinations of the parameters *interval* and *frequency* with values constrained into the repeated operation mode's train set volume of parameters, which consists of *interval*'s values ranging within $[300, 1500]$ and *frequency*'s values ranging within $[7, 120]$. The grey squares found in each figure represent combinations of parameters that lead to performances that could not have been achievable by the inclinometer device during the repeated operation mode's train set acquirement (Subsection 3.6.1), and therefore, the possible predictions of the performances that are driven by parameters' combinations that lie in the grey square are disregarded.

Assessing the characterization of Figure 31, one may observe that the ensemble model could identify three main regions of ASSD values, portrayed as with a colour scale. There is the blue region, which stands for reasonably high ASSD values, the green region, which stands for

Figure 32 – Full characterization of the repeated operation mode for the sensing y-axis' responses.



Source: Author.

ASSD values ranging within $[0.050, 0.066]$, and the yellow/deep yellow region, which stands for the lowest ASSD values reachable by the sensing x-axis, according to the ensemble regression characterization. It is interesting to note that the ensemble model could identify a sweet spot of parameters' values that lead to an optimum performance, roughly located when *interval* is set within $[1200, 1500]$ and *frequency* is set within $[15, 30]$.

Assessing the Figure 32, one may remark two main scaled coloured regions, namely a green/yellow region and a faded orange region. The contrast of the sensing y-axis' performances against the sensing x-axis' is notorious. The sensing y-axis' response at the volume of parameters' defined for both the sensing axes is totally different. Not just the scale of the predicted ASSD values is distinctly, but the relationship of the parameters' interactions is other. This fact is evinced when comparing the predicted response of both the sensing axes for high values of *frequency*, where the higher the parameter' value, the worse become the ASSD predicted responses for the sensing x-axis, and sightly better become the ASSD predicted response for the sensing y-axis. Such predicted phenomenon was already remarked by the DOE characterization in Subsubsection B.3.2, and is here, in the ML approach, evinced with more details. The ensemble model charged to characterize the sensing y-axis' responses could also remark a sweet spot of parameters' values that leads to an optimum performance of measurement, located roughly when *interval* is set within the values $[1200, 1500]$, and *frequency* is set at any value within $[7, 120]$.

The full characterization provided by the two ensemble models charged to characterize the repeated operation mode's performances, found in the figures 31 and 32, might be deployed to provide the better configuration of the parameters for any application employing the repeated operation mode. For instance, if one is interested to save battery and still achieve the best performance in both the sensing axes, a good recommendation would be to set *frequency* 20 and *interval* to 1500. However, if the sampling rate required is high, a good configuration that would satisfy such requirement and still reach a reasonable good performance is when *interval* is set to 300 and *frequency* is set to 30. Regarding the repeated operation mode's characterization, one may satisfactory rely on the predictions achieved, knowing that the predicted values are pretty close to real values, once that the ML approach could achieve considerable good scores at the Radius of Tolerance metric with tolerance $\eta = 0.010$.

The figures 29, 30, 31 and 32 present the full characterization of the inclinometer's quality of measurements, assessed by the ASSD metric, achieved by the machine learning strategy discussed in this work. The four figures mentioned, along with the figure 50, fully characterize the three operation modes of the inclinometer device. Such figures might be deployed in a practical manner by the team of Large-Scale Metrology in order to better satisfy the team's measurement requirements and performances' goals.

5 CONCLUSIONS

In this work was presented the development and evaluation of a characterization approach based on DOE theory and enhanced with machine learning algorithms, deployed to evaluate an inclinometer smart sensor. The inclinometer device, designed and developed by the Large-Scale Metrology team, achieves different noisy performances in the measurements, according to the measurements' parameters set. Thereby, in order to remark the device's features for any set of parameters, the device's behaviour is fully characterized, so that the Large-Scale Metrology team may define the most suitable combination of parameters for any application. Aiming to characterize the inclinometer's performance, experiments must have been carried out in order to expose the patterns of the sensor's measurements and assess these patterns with the DOE/ML approach developed. Such experiments were accomplished through a systematic methodology that ponders the environmental influences on the sensibility of the sensing mechanism and reckons the presence of outliers as a signal of instability unleashed by the parameters settings configured. The systematic methodology yields a single real-valued value, namely the ASSD metric, that scores the quality of the inclinometer's performance due to the parameters' values configured, where the higher is the ASSD value achieved, the worse was the performance, which means that the measurements had a considerable amount of outliers and high amplitude of noise. In order to dynamically conduct the hundreds of experiments accomplished in this work, an autonomous software infrastructure was developed to perform the communication with the sensor device, and store the measurement received into a database. Every experiment presented in this work was conducted by this infrastructure and carried out through the systematic methodology.

The DOE approach carried out in this work could leverage a partial characterization of the device, so that crucial information could have been remarked and the ML approach could thereby have been so developed, that it considers the particularities and limitations of the inclinometer's behaviour due to different parameters' values, and thus could better and more reliably remark and predict the patterns of such behaviour. The DOE model employed in this work could leverage the machine learning approach providing four main knowledge. First, the ASSD values achieved by the sensor due to different settings of parameters constrained into a delimited volume of parameters' values may be classified into two groups, high ASSD values, which stand for values greater than 0.1, and lower ASSD values, which stand for values below the score of 0.1. Each group has its own behaviours' singularities. Second, the interactions among the parameters are mostly nonlinear. Third, each ASSD value has an uncertainty that varies its magnitude according to the parameters configured. The ASSD values greater than the score of 0.1 have a trend to have a high uncertainty, whereas ASSD values lower than 0.1 normally reach a low uncertainty. And fourth, each sensing axis leads to different patterns of measurement, so

that the same parameter set might lead to a different quality of measurement for each sensing axis.

Relying on the DOE approach's conclusions, the machine learning approach was so developed, that it considers the fact that the different ASSD values achieved by the device, due to the setting of combination of parameters inside the chosen delimited volume of parameters, can be grouped into two groups, each one with its own particularities and level of uncertainty, and determines a different criterion of evaluation of predictions for each group, namely the R^2 score for ASSD values greater than 0.1, and the RT metric when the ASSD values lie lower than 0.1. The ML approach also acknowledges the fact that the responses of each sensing axis follow a pattern of quality of measurement different in each sensing axis for the same parameters' values, and thereby designates a particular built ML model to predict each sensing axis' performance. During the training of the machine learning models, the fact that the inclinometer's parameters interactions are mostly nonlinear also is considered, so that more effort was dedicated to testing the performance of nonlinear learning based algorithms, since these models could probably better remark the patterns of nonlinear parameters' interactions in the training set. The ML approach presented also exploits the DOE theory in the acquirement of the dataset, relying on the full factorial design to collect observations with maximum significance, enabling thereby that a small dataset could have trained the learning algorithms.

Whereas the DOE approach can provide a description of the phenomenon's patterns under study, still it lacks of a decisive tool that actually might state an unobserved state of such a phenomenon, considering the knowledge already acquired by the methodology. The DOE approach may just provide an intuition of the phenomenon's unobserved states. On the other hand, the machine learning approach may provide a full characterization of any phenomenon, however the building of the approach must consider the data's particularities, otherwise the predictions achieved would not comply with the reality of the phenomenon, leading to low accurate predictions. Therefore, the hybrid technique presented in this work is consisted of the implementation of the complementation of what is lacking in both the approaches, the DOE and ML, namely exactly each other.

The full characterization achieved in this work provides a useful intuition of the inclinometers' response operation in the standard mode, and a considered accurate map of response when the device is operating at the repeated. The continuous mode is fully characterized by the DOE approach, and thus, its characterization relies on real observations. It is important to evince that the characterization approached in this work covers just a constrained volume of parameters, so that any other parameter's value that is found outer the constrained volume is uncharacterized, yet the graphics provided by machine learning may suggest a trend of the inclinometer's behaviour for combination of parameters beyond the constrained volume here explored, assessing the devices' predicted performance at the frontiers of the delimited volume of parameters. The full characterization achieved in this work could have been much more accurate, had the nature of the dataset were more behaved (i.e. without outliers) and more training data had

been acquired. However, since a single ASSD value acquired is considerable costly in matters of time and battery, the acquirement of a bigger train set was not affordable.

Although the characterization presented in this work well describes the features of the inclinometer device whose the experiments were deployed in the DOE/ML assessment, the DOE approach conveyed the knowledge that such characterisation is only meaningful for the very device employed in this work. Experiments presented in this work (Appendice B) with another inclinometer device revealed the fact that both the devices diverge the behaviour for some settings of parameters, and thus the patterns predicted for one device may not be applicable for the other device. Such fact must be ascertained in future works in order to characterize the level of divergence of performance of each device and determine whether exists some relationship that might explain why the devices' performance is for some configurations the same, and others different. In the meanwhile, the author suggests that the DOE/ML characterization procedure developed in this work may be applied individually for each inclinometer device, following the same ensemble methodology and base learners deployed in this work, and designing the dataset's acquirement to have the maximum significance as also presented in this work. Although quite laborious, it is important that a high sensitive sensor as the inclinometer developed by the Large-Scale Metrology team is fully characterized, so that the best exploitation may be achieved by the technology projected.

REFERENCES

AGGARWAL, C. **Neural networks and deep learning:** a textbook. Yorktown Heights: Springer International Publishing, 2018.

AGGARWAL, P. et al. **MEMS-based integrated navigation**. Norwood: Artech House Publishers, 2010.

ALLEN, M. **Understanding regression analysis**. USA: Springer, 1997.

ANDREJAŠIČ, M. MEMS accelerometers. *In:* SEMINAR MEMS ACCELEROMETERS, University of ljubljana, **Conference proceeding...** Marec. 2008.

ANTONY, J. **Design of experiments for engineers and scientists**. Oxford: Elsevier, 2014.

BALABIN, R. M.; LOMAKINA, E. I. Support vector machine regression (svr/ls-svm)—an alternative to neural networks (ann) for analytical chemistry? comparison of nonlinear methods on near infrared (nir) spectroscopy data. **Analyst**, v. 136, p. 1703–1712, 2011.

BISGAARD, S. Blocking generators for small 2k-p designs. **Journal of Quality Technology**, v. 26, n. 4, p. 288–296, 1994.

BISHOP, C. **Pattern recognition and machine learning**. USA: Springer, 2006.

BOUSQUET, O.; LUXBURG, U. v.; RATSCH, G. **Advanced lectures on machine learning**. Berlin: SpringerVerlag, 2004.

CHAI, T.; DRAXLER, R. R. Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature. **Geoscientific Model Development**, v. 7, n. 3, p. 1247–1250, 2014.

CHEN, T.; GUESTRIN, C. Xgboost: a scalable tree boosting system. **Association for Computing Machinery**, p. 785–794, 2016.

CYPRESS. **CYBLE-212006-01**. 2019. Available in: https://www.cypress.com/file/318881/download. Accessed in 26 aug. 2020.

DAHLEM, P. et al. Hybrid model approaches for compensating environmental influences in machine tools using integrated sensors. **at - Automatisierungstechnik**, De Gruyter, Berlin, Boston, v. 68, n. 6, p. 465 – 476, 2020.

DAI, R. et al. Application of tilt sensors in functional electrical stimulation. **IEEE Transactions on Rehabilitation Engineering**, v. 4, n. 2, p. 63–72, 1996.

DEAN, A.; VOSS, D.; DRAGULJIĆ, D. **Design and Analysis of Experiments**. Switzerland: Springer International Publishing, 2017.

FABER, F. A. et al. Machine learning energies of 2 million elpasolite(abc2d6)crystals. **Physical Review Letters**, American Physical Society (APS), v. 117, n. 13, Sep 2016.

FRANK, R. **Understanding smart sensors**. Norwood: Artech House, 2013.

GOODFELLOW et al. **Deep learning**. Massachusetts: The MIT Press, 2016.

GRAFANA. **Grafana dashboard**. 2020. Avalaible in: https://grafana.com/. Accessed in 1 dez. 2020.

GUPTA, N. **Inside bluetooth low energy**. USA: Artech House, 2013.

HÄRDLE, W. K.; SIMAR, L. **Applied multivariate statistical analysis**. Berlin: Springer-Verlag, 2015.

INFLUXDATA. **InfluxDB**. 2020. Avalaible in: https://www.influxdata.com/products/influxdb/. Accessed in 30 nov. 2020.

KAMATH, C.; FAN, Y. J. Regression with small data sets: a case study using code surrogates in additive manufacturing. **Knowledge and Information Systems**, v. 57, n. 2, p. 475–493, 2018.

KRAMER, O. **Dimensionality reduction with unsupervised nearest neighbors**. Berlin: [s.n.], 2013.

LIGHT, R. A. Mosquitto: server and client implementation of the mqtt protocol. **Journal of Open Source Software**, v. 2, n. 13, p. 265, 2017.

LUCKNER, M.; TOPOLSKI, B.; MAZUREK, M. Application of xgboost algorithm in fingerprinting localisation task. **Springer International Publishing**, 2017.

LUU, K. et al. Age estimation using active appearance models and support vector machine regression. **2009 IEEE 3rd International Conference on Biometrics**, p. 1–5, 2009.

LYSHEVSKI, S. E. **MEMS and NEMS:** systems, devices, and structures. USA: CRC Press, 2002.

MOHD-YASIN, F.; KORMAN, C. E.; NAGEL, D. J. Measurement of noise characteristics of mems accelerometers. **Solid-State Electronics**, v. 47, n. 2, p. 357 – 360, 2003.

MQTT. **MQTT: The Standard for IoT Messaging**. 2020. Avalaible in: https://mqtt.org/. Accessed in 30 nov. 2020.

MURATA. **Murata's devices**. 2020. Available in: https://cutt.ly/VztP6gG. Accessed in 06 set. 2020.

MURATA. **SCA3300**. 2020. Avalaible in: https://www.murata.com/-/media/webrenewal/products/sensor/pdf/datasheet/datasheet_sca3300-d01.ashx?la=en-gb&cvid=20190620010315610400. Accessed in 23 nov. 2020.

MURATA. **SCA830**. 2020. Available in: https://www.murata.com/-/media/webrenewal/products/sensor/pdf/datasheet/datasheet_sca830-d07.ashx?la=en-gb&cvid=20190118045358805600. Accessed in 26 aug. 2020.

MURPHY, K. P. **Machine learning:** A Probabilistic Perspective. Massachusetts: The MIT Press, 2012.

OSBORNE, J.; OVERBAY, A. The power of outliers (and why researchers should always check for them). **Pract. Assess. Res. Eval.**, v. 9, 01 2004.

PARTNERSHIP, P. F. **An introduction to MEMS (Micro-electromechanical Systems)**. Loughborough: Faraday Partnership Wolfson School of Mechanical and Manufacturing Engineering Loughborough University, 2002.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

SCHMIDT, J. et al. Predicting the thermodynamic stability of solids combining density functional theory and machine learning. **Chemistry of Materials**, American Chemical Society, v. 29, n. 12, p. 5090–5103, Jun 2017.

SKIENA, S. S. **The data science design manual**. USA: Springer Publishing Company, Incorporated, 2017.

SU, G. et al. Gaussian process machine-learning method for structural reliability analysis. **Advances in Structural Engineering**, v. 17, n. 9, p. 1257–1270, 2014.

TANAKA, S.; NISHIFUJI, S. Automatic on-line measurement of ship's attitude by use of a servo-type accelerometer and inclinometers. **IEEE Transactions on Instrumentation and Measurement**, v. 45, n. 1, p. 209–217, 1996.

TANENBAUM, A. S.; BOS, H. **Modern Operating Systems**. USA: Prentice Hall Press, 2014.

TOWNSEND, K.; CUFÍ, C.; DAVIDSON, R. **Getting started with bluetooth low energy:** tools and techniques for low-power networking. USA: O'Reilly Media, 2014.

WAUTERS, M.; VANHOUCKE, M. Support vector machine regression for project control forecasting. **Automation in Construction**, v. 47, p. 92 – 106, 2014.

WEBB, G. I.; ZHENG, Z. Multistrategy ensemble learning: reducing error by combining ensemble learning techniques. **IEEE Transactions on Knowledge and Data Engineering**, v. 16, n. 8, p. 980–991, 2004.

YATES, F. Sir ronald fisher and the design of experiments. **Biometrics**, [Wiley, International Biometric Society], v. 20, n. 2, p. 307–321, 1964. Disponível em: http://www.jstor.org/stable/2528399.

ZHANG, Y.; LING, C. A strategy to apply machine learning to small datasets in materials science. **npj Computational Materials**, v. 4, n. 1, p. 25, May 2018.

ZHOU, Z. **Ensemble methods:** foundations and algorithms. USA: CRC Press, 2012.

A APPENDICE - INCLINOMETER'S OPERATION MODE ALGORITHMS

In this appendice is presented pseudocode algorithms that describes the logistic of the inclinometer's operation modes employment.

Algorithm 2: Procedure for the STANDARD operation mode

```
1 initialize_bluetooth_connection();
2 while bluetooth_connected do
3 |   incl_x, incl_y, std_x, std_y = read_STANDARD_mode( interval, frequency, nValue);
4 |   timesleep( seconds );
5 end
```

Algorithm 3: Procedure for the REPEATED operation mode

```
1 initialize_bluetooth_connection();
2 initialize_REPEATED_mode( interval, frequency, nValue );
3 while bluetooth_connected do
4 |   if new_data_received();
5 |    then  incl_x, incl_y, std_x, std_y = new_data;
6 |
7 |   end
8 end
```

Algorithm 4: Procedure for the CONTINUOUS operation mode

```
1 initialize_bluetooth_connection();
2 initialize_CONTINUOUS_mode( frequency, nValue );
3 while bluetooth_connected do
4 |   if new_data_received();
5 |    then incl_x, incl_y = new_data;
6 |
7 |   end
8 end
```

B APPENDICE - INCLINOMETER'S DOE PERFORMANCE

In this appendice is presented the formulation and driving of the DOE model structured to partially characterise the features of the inclinometer, so that an intuition of the device's behaviour might be accomplished.

B.1 REPLICATION, BLOCKING AND RANDOMIZATION IN THE INCLINOMETER'S EXPERIMENT DESIGN

As explained in Subsection 2.3.1, there exist three foundations concerning the planning and design of experiments, namely Replication, Blocking and Randomization. The DOE experiments in this work were so architectured that their designs comply with these three principles.

In this work, DOE was employed so that a conjunct of trials could define a delimited volume of combinations of parameters. A conjunct of trials performs the experimentation of all the possible combinations of a given list of parameters, following the structure of a full factorial design. For instance, given the parameter's values of Table 13, a conjunct of trials that performs all the possible combinations for the given factors' levels is structured as a full factorial design, sketched in Table 14. Since just three parameters are varying at two levels, it is said that this approach is a 3^2 full factorial design.

Table 13 – List of parameter's values of the conjunct of trials of Table 14.

Parameter	**Symbol**	**Minimum Value**	**Maximum Value**
Interval	In	300	1500
Frequency	Fr	5	90
Timesleep	Ts	0.4	1.3
nValue	nV	10	10

Table 14 – Conjunct of trials of the given parameters of Table 13.

Trial	**In**	**Fr**	**Ts**	**nV**
1	+1	+1	+1	-
2	+1	+1	-1	-
3	+1	-1	+1	-
4	+1	-1	-1	-
5	-1	+1	+1	-
6	-1	+1	-1	-
7	-1	-1	+1	-
8	-1	-1	-1	-

In the DOE experiments of this work, whenever a list of parameters, as the in Table 13, is structured to be carried out, its respective conjunct of trials is performed three times, in an interval of three or four days. By the means of the arrangement of executing the conjunct of trials three times, during different days, the principle of blocking is being obeyed. By sorting several executions of the same conjunct of trials within days, different sources of variability like day-to-day can be noted in the conjunct of trials' responses, and the average of the respective trials' response can provide a more accurate estimation of the device's response for the parameters settled in that trial.

Regarding the performance of the conjunct of trials, in the experiments of this work the trials' execution order is randomized. In each one of the three times executed conjunct of trials of a parameters' list, the sequential order of execution of each trial in one conjunct of trials is different from the others conjuncts of trials. This randomization of the trials is autonomously driven by the infrastructure developed, depicted in Section 3.2. The proper trial's randomization order of execution comply with the DOE principle of randomization. This strategy is established in order to avoid experimental biases, and achieves such aim by equally distributing noise elements among different levels of a factor, and thereby the effect of noisy singularities is averaged through the performance of the conjunct of trials.

The last principle to be architected into the inclinometer's DOE is replication. For the inclinometer's DOE scenario, and considering that the experiments of this work were carried out during a lockdown due to COVID-19 pandemic, the easier and properer formulation of applying replication is to replicate the performance of some conjunct of trials with different inclinometers devices, and then average the respective trials' responses from each conjunct of each device. This design was planned and executed, but immediately results conveyed the information that the two inclinometer devices, which the author had at the moment of the experiments, had some serious divergences regarding the response for the same set of parameters. These divergences are properly approached in Subsection B.4. However, such divergences had led the decision of to do not include replication considering different devices into the inclinometer's DOE, since the average of the response for the same parameters setting in both the devices would lead to an unreliable value for both the sensors.

B.2 INCLINOMETER'S DESIGN OF EXPERIMENTS

As previously presented in Subsection 3.1.1, the inclinometer device performs according to some parameters that drive the measurement's performance. These parameters are configured at the moment of the Bluetooth connection, by the means of the infrastructure developed and detailed in Subsection 3.2. The inclinometer's design of experiments is so architected that by varying the levels of the device's parameters, a variation of the measurement's performance and quality is remarked. In the following DOE, the inclinometer's performance is assessed through the systematic methodology explained in Section 3.4. Since the noise component of

the measurements is the main aspect that influences the sensor's accuracy, performance, and reliability, the output response in interest, and evaluated in the inclinometer's DOE of this work, is the Average of the Segments' Standard Deviation (ASSD).

As explained beforehand, the inclinometer device has five configurable measurement parameters, namely *interval*, *frequency*, *nValue*, *operation mode* and, for the standard operation mode, *timesleep* (Subsection 3.1.1). The first step to architect a design of experiments is to identify the controllable and uncontrollable factors of the experiment under study. For the inclinometer device, the five configurable measurement parameters are assigned as controllable variables. The environment's influence and unknown sources of measurement's variability are assigned, in this work, as uncontrollable variables. Since just controllable variables can be manipulated, the inclinometer's DOE is characterised by the five configurable measurement parameters.

In order to assess the significance of the varying of a parameter in the ASSD metric, immediately experiments, which do not belong to the DOE approach, were carried out. The result of these first experiments conveyed the knowledge that the changing of value of the parameter nValue does not lead to a meaningful difference in the device response. In fact, the routine in which the nValue acts is a recommended strategy from the manufactures of the MEMS accelerometer Murata (2020b) employed in the inclinometer device. The manufactures assert that exists the possibility that, when required to read the acceleration reported by the sensing mechanism, the sensor may fail to accomplish the measurement. By the setting of a routine of reading the sensing mechanism for 'nValue' times, the chance of a faulty response is minimized. Besides, an averaged measurement of the sensing mechanism is a more reliable response. Thus, since the varying of *nValue* does not lead to a significant change in the measurement response, all the following experiments and designs approached in this work consider a fixed value for the parameter *nValue*, such that *nValue* is equal to 10.

Thereby, the remnant configurable parameters compose the inclinometer's DOE. As presented in Subsection 3.1.1, the device's measurement routine is dictated by the operation mode. Each operation mode considers a certain set of parameters in its performance. The standard mode employs the parameters *interval*, *frequency*, *timesleep* and *nValue*. The repeated mode employs the same parameters as the standard operation mode, but the parameter *timesleep* does not exist for this procedure. And the last operation mode, continuous, employs just the parameters *frequency* and *nValue*. The different behaviours and parameters involved in the routine of each operation mode leads to the conclusion that each operation mode must have its own DOE. Thus, for each *operation mode*, a different conjunct of trials was designed and settled.

In Subsection 3.4 was presented the systematic methodology, and it was shown as example experiments that were left laid still on an assumed plain surface collecting acceleration data for 60 minutes, where the first five minutes are a standstill time in which the device does nothing, and whose the first six minutes of collected data are disregarded. This exact methodology is followed for each of the design further presented, however for each operation mode, a specific

time of collecting data is assigned to an experiment. Each experiment belonging to the standard mode is performed for 60 minutes (49 minutes when disregarding the time of standstill and initial offset), and each experiment belonged to the repeated and continuous mode were executed for 25 minutes (14 minutes when disregarding the time of standstill and initial offset). Since that in the repeated and continuous mode the parameter *timesleep*, which delays the measurement, is not applied, in these modes the rate of sampling is greater than in the standard mode. Thus, in order to have roughly the same amount of samples in an experiment belonged to any operation mode, the time of execution of an experiment in the repeated and continuous mode is smaller than an experiment belonging to the standard mode.

The standard operation mode is composed by four parameters, and thus is the most complex mode. For this mode of operation, the delimited volume of parameters is bounded by the list of parameters presented in Table 15. The respective conjunct of trials' design is conveyed in Table 16. Each conjunct of trials was performed three times, and the sequential order of execution of each trial, in each conjunct of trials, was random. All the experiments were performed through the exact same systematic methodology presented in Subsection 3.4. Since a conjunct of trials belonging to this mode lasts eight hours to become fulfilled, for each night of experiments, one execution of a conjunct of trials was performed.

Table 15 – List of parameter's values of the first step of the conjunct of trials from the standard operation mode, whose tables 16 depicts its design.

Parameter	**Symbol**	**Minimum Value**	**Maximum Value**
Interval	In	500	1200
Frequency	Fr	5	120
Timesleep	Ts	0.4	1.3
nValue	nV	10	10

Table 16 – Conjunct of trials of the given parameters of Table 15, 17 and 18.

Trial	**In**	**Fr**	**Ts**	**nV**
1	+1	+1	+1	-
2	+1	+1	-1	-
3	+1	-1	+1	-
4	+1	-1	-1	-
5	-1	+1	+1	-
6	-1	+1	-1	-
7	-1	-1	+1	-
8	-1	-1	-1	-

In order to have a better understanding of the device's behaviour for settings of parameters inner the bounded volume of parameters from Table 15, two other different conjuncts of trials were carried out for the standard mode. The goal of the driving of two more different

conjuncts of trials is to follow the DOE methodology in order to find a sweet spot of factor's combination levels that leads to the smallest ASSD value. Thereby, here is introduced another terminology to organize the sequence of an inner conjunct of trials. The first volume of parameters is named first step, and is assigned to the standard mode through the Table 15. The second volume of parameters is bounded inner the first step, and is appointed as second step. The list of parameters for the second step of the standard mode is presented in Table 17, and the respective conjunct of trials design is portrayed in Table 16.

Table 17 – List of parameter's values of the second step of the conjunct of trials from the standard operation mode, whose Table 16 depicts its design.

Parameter	**Symbol**	**Minimum Value**	**Maximum Value**
Interval	In	850	1200
Frequency	Fr	63	120
Timesleep	Ts	0.85	1.3
nValue	nV	10	10

Finally, the third volume of parameters is bounded inner the second step volume, and is named as third step. Its list of parameters can be remarked in Table 18 and its respective conjunct of trials design in Table 16.

Table 18 – List of parameter's values of the third step of the conjunct of trials of the standard operation mode, whose Table 16 depicts its design.

Parameter	**Symbol**	**Minimum Value**	**Maximum Value**
Interval	In	1025	1200
Frequency	Fr	63	91
Timesleep	Ts	0.85	1.07
nValue	nV	10	10

The choice of the parameters that bound the volume of parameters of the first step was so decide because the Large-Scale Metrology's team was interested in constantly and shortly spaced in time data, which justifies the choice of the scale of values for the parameters *interval* and *timesleep*. The parameter *frequency* was so settled that it respects the limitations of the MEMS accelerometer IC, stated in its datasheet (MURATA, 2020b) and attested empirically. It was verified that, for the standard mode, frequencies greater than 120 Hz and smaller than 5 Hz may lead to unstable and unreliable behaviours. Thus, the parameter *frequency* was bounded as conveys Table 15. The parameters from the second and third conjunct of trials were so determined that they follow the gradient of better performance from the previous step, in order to find a sweet spot for the smaller value of the ASSD metric. This gradient of better performance and the choice of the parameters' values of the second and third step are approached and discussed again with more details in Subsection B.3.

The repeated operation mode consists of three parameters. Due to its measurement's

routine configuration, the existence of the *timesleep* parameter is not applicable. Thus, the repeated mode of operation's DOE is so architected that just two parameters vary. The repeated operation mode's DOE is also structured with inner steps of a delimited volume of parameters. In the first step, the boundaries of the volume of parameters are delimited and presented in Table 19. The Table 20 portrays the design of the experiments for this first step. It is important to remark that, since just two parameters are varying in this mode at two different levels, its design is labelled as a 2^2 full factorial design. All the three steps of experiments belonging to the repeated operation mode were performed similarly as they were for the standard mode, repeating each conjunct of trials of each step three times and following the systematic methodology depicted in Subsection 3.4. Since that a conjunct of trials of the repeated mode lasts one hour and forty minutes to become fulfilled, roughly two executions of conjuncts of trials were performed in one night of experiments.

Table 19 – List of parameter's values of the first step of the conjunct of trials from the repeated operation mode, whose Table 20 depicts its design.

Parameter	**Symbol**	**Minimum Value**	**Maximum Value**
Interval	In	400	1400
Frequency	Fr	7	120
nValue	nV	10	10

Table 20 – Conjunct of trials of the given parameters of Table 19, 21 and 22.

Trial	**In**	**Fr**	**nV**
1	+1	+1	-
2	+1	-1	-
3	-1	-1	-
4	-1	-1	**-**

The second inner step of the repeated operation mode's volume of parameters is settled in order to find the sweet spot among the parameter's levels combination, so that the ASSD metric has the smallest value. Its boundaries are presented in Table 21 and its respective conjunct of trials portrayed in Table 20.

Table 21 – List of parameter's values of the second step of the conjunct of trials from the repeated operation mode, whose Table 20 depicts its design.

Parameter	**Symbol**	**Minimum Value**	**Maximum Value**
Interval	In	900	1400
Frequency	Fr	63	120
nValue	nV	10	10

The volume of the second step was also explored by a third step of inner parameter's boundaries. The parameters of this third step can be found in Table 22, and its respective conjunct

of trials design is portrayed in Table 20.

Table 22 – List of parameter's values of the third step of the conjunct of trials of the repeated operation mode, whose Table 20 depicts its design.

Parameter	Symbol	Minimum Value	Maximum Value
Interval	In	1150	1400
Frequency	Fr	63	91
nValue	nV	10	10

The choice of the parameters that consist the first step of the conjunct of trials from the repeated operation mode is justified by the interests of the Large-Scale Metrology's team. Since the team was interested in data shortly spaced in time, the values of *interval* were so settled as conveys Table 19. The maximum value of *frequency* for the first step was so settled that it complies with the device's capacity presented in its datasheet (MURATA, 2020b), and respects the limitations empirically observed. It was also remarked that, for the repeated mode, frequencies smaller than 7 lead to unstable and unreliable behaviour. The choice of parameters for the second and third inner steps was so determined that they follow the gradient of smaller ASSD value. More details about the choice of the inner parameters are approached in Subsection B.3.

Finally, the last DOE architected was structured for the continuous operation mode. This routine works through the setting of just two parameters, namely *frequency* and *nValue*. Since that in this work *nValue* has a fixed value, the experiments regarding the continuous mode are driven by the varying of just one parameter, the *frequency*. For the changing of just one parameter, the DOE model is simplified to the simple performance of trials in the specified levels of the factor under analyses. For this situation, there is no need of a full factorial design. Thus, the DOE for the continuous mode is consisted of one single step, with a respective conjunct of trials that is executed three times in random trial sequences through the systematic methodology. The Table 23 depicts the parameter's configuration of the experiments of the continuous operation mode, and the Table 24 portrays its conjunct of trials.

Table 23 – List of parameter's values of the first and unique step of the conjunct of trials of the continuous operation mode, whose Table 24 depicts its design.

Parameter	Symbol	Value 1	Value 2	Value 3	Value 4
Frequency	Fr	1	2	3	4
nValue	nV	10	10	10	10

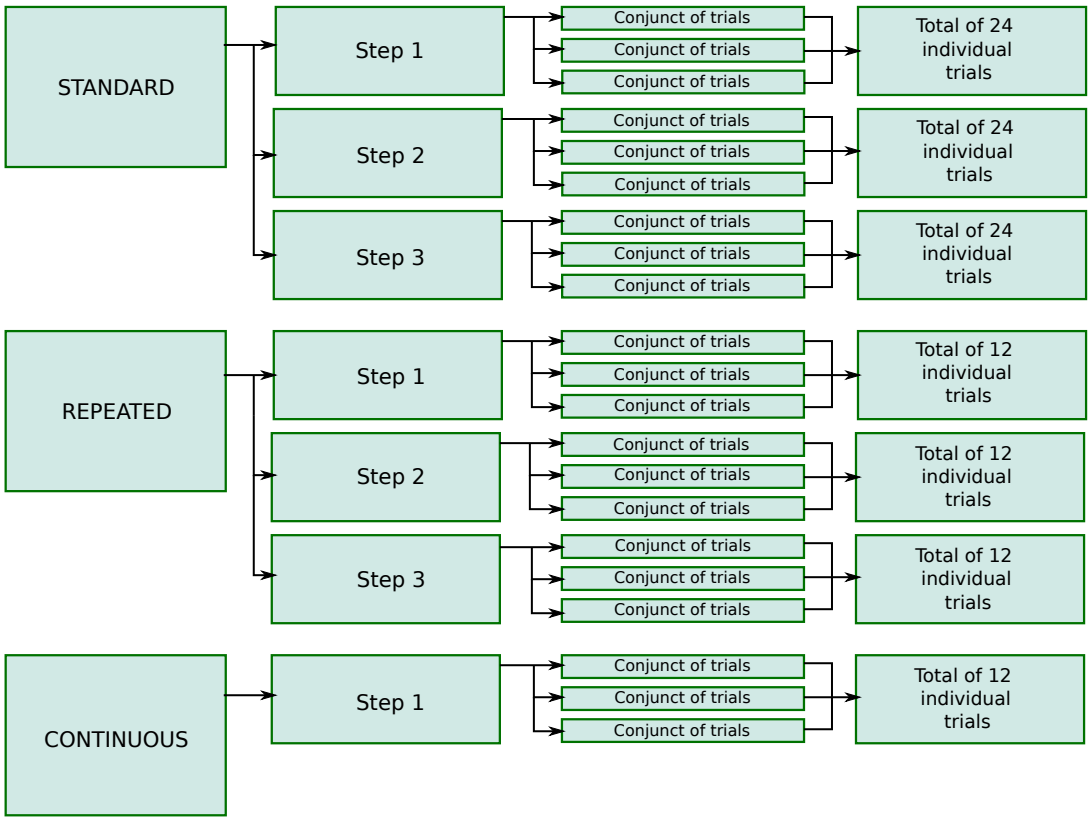
The choice of the values that consist the parameter *frequency* was determined in order to comply with the Large-Scale Metrology's team objectives. Besides, it was empirically observed that, for the continuous mode, frequencies greater than 4 Hz leads to an unstable scenario. The limitation in this case is so that the Bluetooth protocol can not unrelenting operate at the same rate as the requirements and responses in frequencies greater than 4 Hz.

Table 24 – Conjunct of trials of the given parameters of Table 23.

Trial	Fr	nV
1	Value 1	-
2	Value 2	-
3	Value 3	-
4	Value 4	-

In order to summarize the schematic of the inclinometer's DOE, the Figure 33 presents the organization of performances among operation mode, step, conjunct of trials and trial.

Figure 33 – Schematic portraying the arrangement of the experiments that consist the inclinometer's DOE.



Source: Author.

B.3 INCLINOMETER'S DOE ASSESSMENT

In this subsection are presented the performances of the DOE experiments, regarding each operation mode, as depicts Figure 33. The conclusions of the DOE analyses are further the foundations of the development of the machine learning approach, found in Section 3.6. Subsequently is presented the assessment of the inclinometer's response for each operation mode, for a given parameters' configuration. Cube plots and interaction plots are here employed to

convey and enlighten the characteristics of the inclinometer's behaviour. The ASSD metric is employed to assess the inclinometer's features. All the preprocessing of the inclinometer's data was performed as dictates the systematic methodology in Section 3.4.

B.3.1 Stantard operation mode DOE

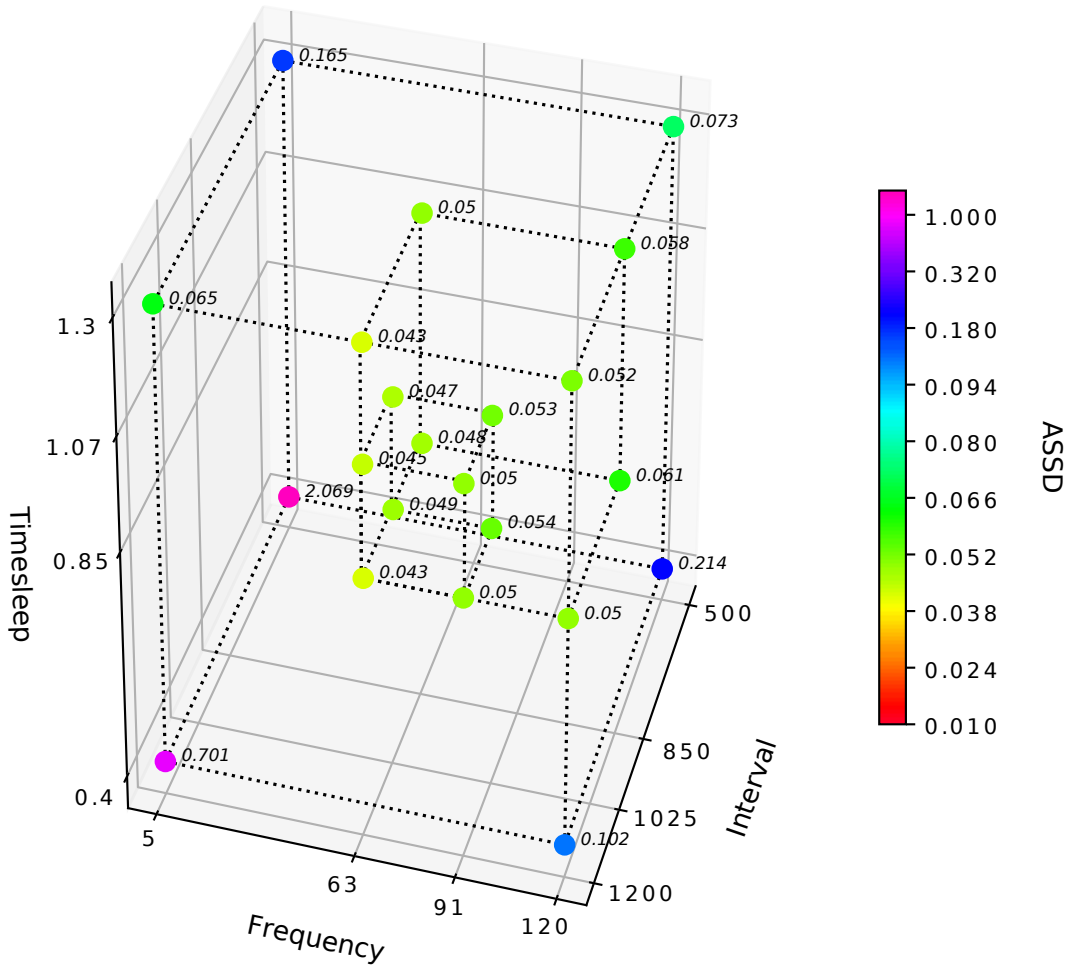
The DOE for the standard operation mode is consisted of varying three configurable parameters, namely *interval*, *frequency* and *timesleep*. The designs of the standard operation mode are structured into steps, as conveys Figure 33, according to the respective conjunct of trials presented in tables 15, 17 and 18. In order to easily convey and assess the response of this operation mode's DOE, cube plots and interactions plots (Subsection 2.3.3) were deployed to evaluate each sensing axis.

The figures 34 and 35 present the full cube plot of all the three steps. Figure 34 conveys the response of the ASSD metric for the sensing x-axis, whereas Figure 35 stands for the sensing y-axis. In these cube plots, one may easily remark the structure of the steps into a volume of parameters. The most outer delimited cube stands for the first step, whereas the inner cubes stand for the second and third step, where the smallest and innermost cube represents the third step. A colour scale was employed to easier identify the gradient of better response for the ASSD metric, where, the closer to zero is the ASSD value, the more stable and smaller the noise component of the response whose the parameters combination had led to. Each coloured point represents the average of the three trials' ASSD values, which refers to the performance of the three times executed conjunct of trials belonging to the step in analysis. Since that in the cube plots, for each inner volume of parameters, one set of parameters (i.e. one coloured point) is shared by two cube plots, whenever this graphic particularity occurs, the response of the most outer volume of parameters is displayed. The real value of the responses that were not exhibited can be found in the tables 26 and 27.

The assessment of the standard operation mode's DOE begins following the execution order of the steps, which was sequential from step one until step three. As formerly mentioned, the outermost delimited cube stands for the first step. Assessing the first step, the Figure 34 conveys the information that, for the sensing x-axis, low levels of *interval*, *frequency* and *timesleep* lead to a pretty unstable behaviour. However, the most meaningful parameter that indeed changed the scale of the ASSD metric was *timesleep*. Still in Figure 34 assessing the outermost cube, one may remark clearly that when *timesleep* is set to a high level, the ASSD responses are drastically reduced. Similarly, in this first step of the sensing y-axis, Figure 35 conveys responses pretty akin to the respective set of parameters of the sensing x-axis, where the lower the level of the combination of parameters, the greater the value of the ASSD metric, and thereby worse the measurement's performance. The values of the ASSD metric at the same combination of parameters for both the sensing axes in this first step are quite close. Nevertheless, one may remark that the performances for the sensing y-axis were better for every trial of this first step.

An assessment of the variability of the conjunct of trials' performance can be conveyed

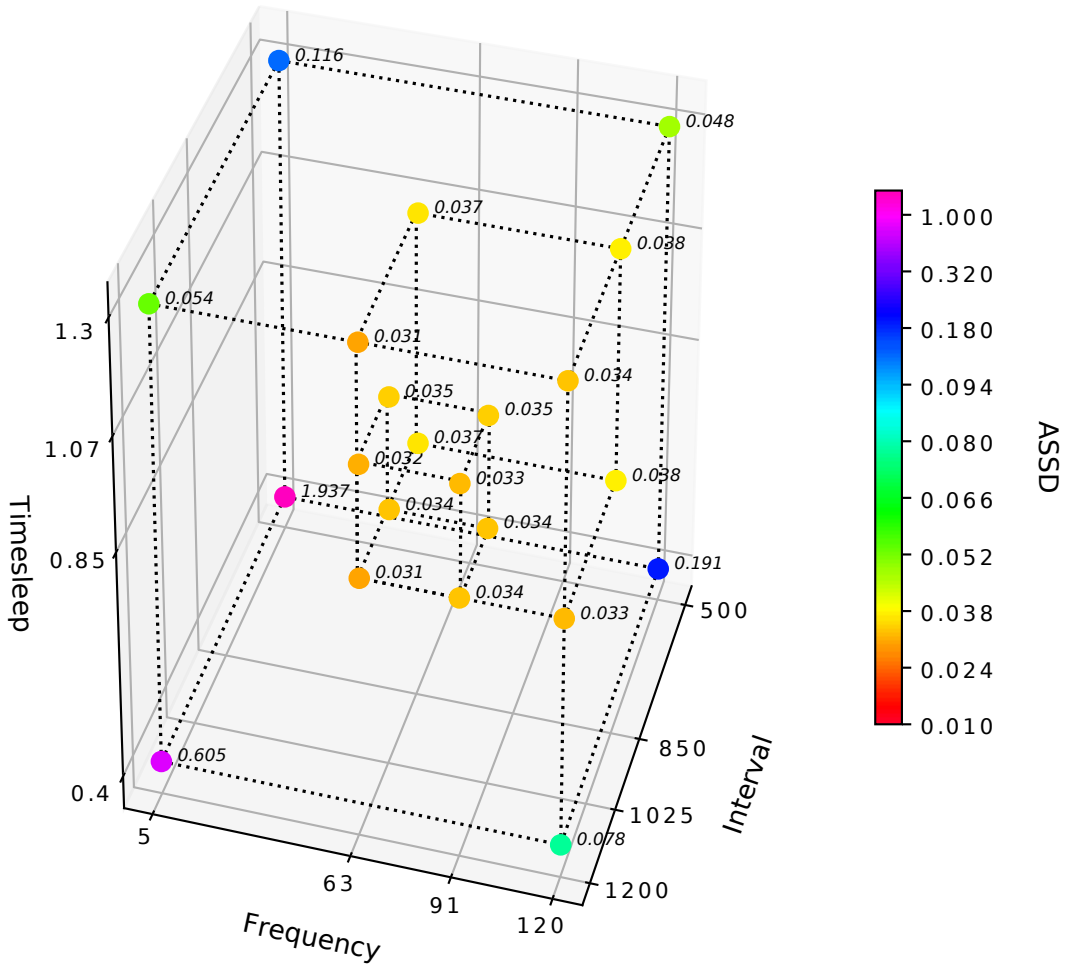
Figure 34 – Cube plot of the three steps of experiments from the standard operation mode, representing the response of the ASSD metric for the sensing x-axis. The most outer delimited cube stands for the first step, the middle inner cube for the second step, and the smallest and innermost cube for the third step. One may remark that, the closer to zero is the ASSD metric value, the more stable and smaller is the noise component of the response whose a combination of parameters leads.



Source: Author.

by the means of the graphics from Figure 36 and Table 25, where the trials from each conjunct of trials belonging to the first step, whose the combination of parameters is presented in Table 15 and its trials' design depicted in Table 16, is plotted in Fig.36 left, for the sensing x-axis, and Fig.36 right for the sensing y-axis. The information conveyed by the Table 25 is more clearly remarked in Figure 36, where the reliability of the systematic methodology is attested, since all the respective trials from each conjunct of trials present ASSD values in the same scale. In fact, this scale can be quantified by the standard deviation of the respective trials from the three

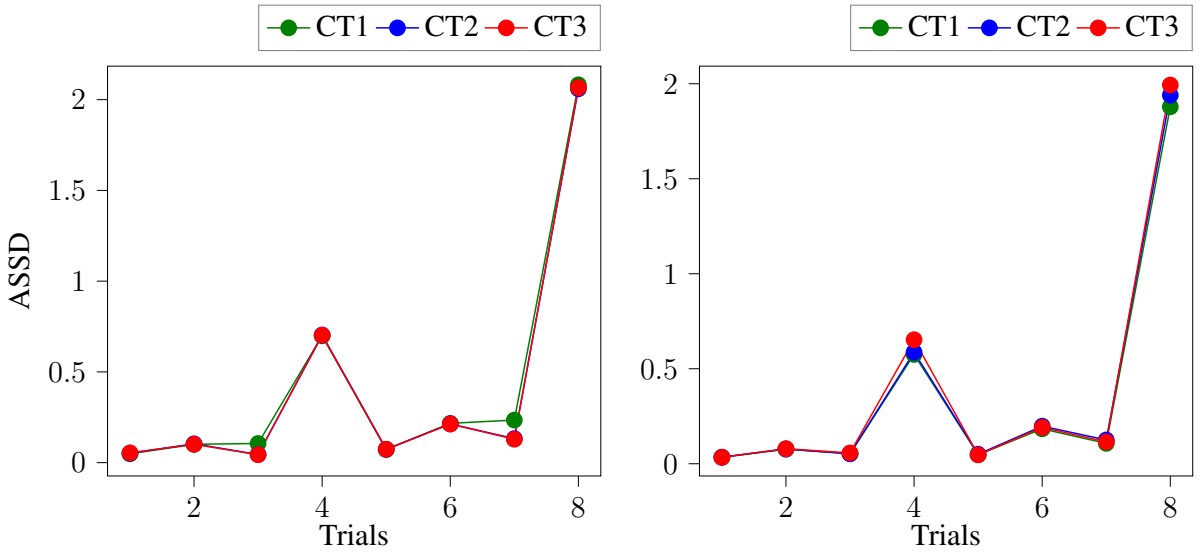
Figure 35 – Cube plot of the three steps of experiments from the standard operation mode, representing the response of the ASSD metric for the sensing y-axis. The most outer delimited cube stands for the first step, the middle inner cube for the second step, and the smallest and innermost cube for the third step. One may remark that, the closer to zero is the ASSD metric value, the more stable and smaller is the noise component of the response whose a combination of parameters leads.



Source: Author.

conjunct of trials, as presented in Table 25 by the abbreviation 'Std', followed by the respective sensing axis' trial mean ASSD value. One may observe that, almost all the standard deviations values, for both the sensing axis, are at least tenfold smaller than the respective mean value. There are two facts observed in Figure 36 and Table 25 that are further established as a foundation on which relies the machine learning approach for the standard operation mode. First, is attested that the performance of the sensor for a given combination of parameters keeps the same behaviour. Each different set of parameters leads to a unique systematic pattern of response, which repeats

Figure 36 – Graphics presenting the results of the ASSD metric for all the three conjuncts of trials performed in the first step. The number of trials, found in the abscissa axis stands for the numeration of the trial's design from Table 16. The graphic from figure left regards the response for the sensing x-axis, whereas the graphic in the right concerns for the sensing y-axis. The abbreviation 'CT' refers to Conjunct of Trials.



Source: Author.

whenever the respective set of parameters is settled. Second, the systematic methodology could attest roughly the same ASSD value for each respective trial from the three conjunct of trials. The relevance of these facts for the development of the machine learning conceptualization is further approached in Section 3.6.

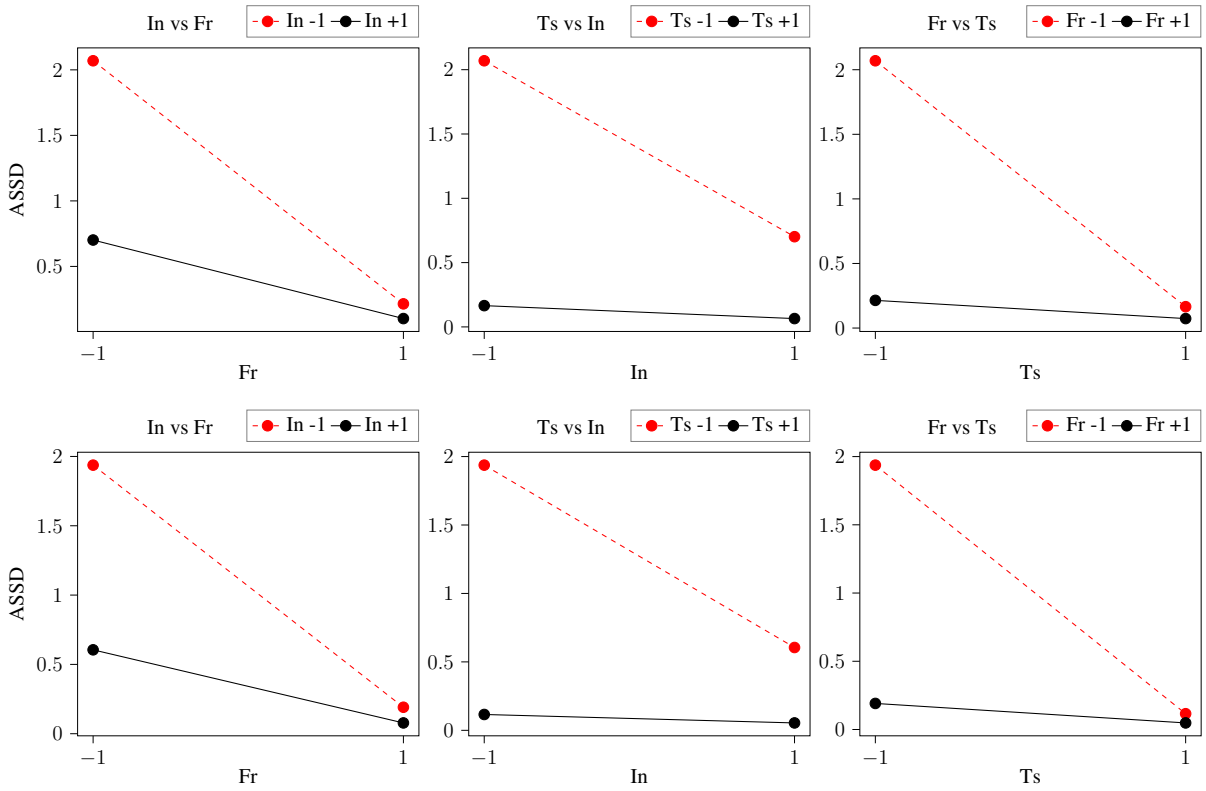
In order to evaluate the interactions of the parameters with themselves, for this first step of the standard operation mode's DOE, interactions plots (Subsection 2.3.3) were arranged for each sensing axis, presented in Figure 37. The first horizontal row of graphics in Fig.37 conveys the interactions among the parameters for the sensing x-axis' response, whereas the row below stands for the response of the sensing y-axis. For every interaction plot presented in this work, the remnant parameter of one interaction plot is set to low level. For instance, the first interaction plot of the first row of figure 37 presents the interaction between the parameters *interval* and *frequency*, assessing the ASSD response when these factors vary their levels from low to high level. For the scenario described, the remnant factor *timesleep* is set to low level, whereas the factors *interval* and *frequency* vary.

The interactions plots have an intuitive propriety that allow one to assess the features of the influence that the changing of one parameter's value may affect the influence of the other parameters in the final response under analyses. When the lines of one interactions plot, presented in colours red and black in this work, are parallel, there is no interaction between the

Table 25 – Table presenting the mean and standard deviation value, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the first step belonged to the standard operation mode.

Trial number	ASSD mean x-axis	Std x-axis	ASSD mean y-axis	Std y-axis
1	0.0515	0.0026	0.0341	0.0004
2	0.1020	0.0016	0.0778	0.0012
3	0.0647	0.0355	0.0540	0.0024
4	0.7014	0.0022	0.6050	0.0418
5	0.0733	0.0009	0.0481	0.0018
6	0.2142	0.0020	0.1909	0.0070
7	0.1653	0.0600	0.1157	0.0090
8	2.0692	0.0114	1.9373	0.0573

Figure 37 – Interactions plots of the first step of experiments from the standard operation mode. The first row of group of three plots stands for the ASSD values of the sensing x-axis' performance, whereas the second row stands for the sensing y-axis' performance. For each interaction plot, the remnant parameter is set to low level.



Source: Author.

parameters of the plot under analysis. When the degree of non-parallelism between the lines is subtle, there is interaction between the parameters presented, but it is not considerable. However,

when the degree of non-parallelism between the lines is quite notable, one may assert that there is a considerable interaction between the parameters under analyses. Thus, evaluating the first row of Figure 37, one may remark that, since the degree of non-parallelism of all those plots of the first row is quite huge, it is said that there is a notable interaction among the parameters. In fact, the interactions between *interval* and *frequency*, and *frequency* and *timesleep*, seems to have roughly the same degree of non-parallelism. The interaction between *timesleep* and *interval*, although notably attested, seems to be less considerable than the interaction for the other two arrangements, since the degree of non-parallelism presented between the lines is smaller. Evaluating the second row of the interaction plots from Figure 37, which stands for the sensing y-axis' ASSD response, one may visually conclude that the degree of non-parallelism of each plot is almost the same as its respective plot in the first row, and thus, the same observations remarked for the sensing x-axis are applicable for the sensing y-axis.

The conclusion that the knowledge of the existence of interaction among all the factors of the standard mode leads, is that non-linearity rules the behaviour of the combination of parameters effect, regarding the assessment of the ASSD metric. The increase or decrease of one factor's level will affect the effect that the level of other factors lead, and thus, all the factors in this operation mode contribute with equally high influence on the performance of the final response under analyses, since the smallest variation of any parameter's value unleashes disturbances in the contribution of the other parameters. This scenario conveys that, to achieve a sweet spot of combination of parameters, in order to reach better performance in the ASSD metric, is not a trivial endeavour. To accomplish such an aim, one would need to carry out more experiments with combinations of parameters that follow the gradient of better response.

In order to comply with the aim of to find a sweet spot of combinations of parameters that lead to a performance that reaches the smaller ASSD values, the standard operation mode's first step of experiments is explored, so that combination of parameters inner the boundaries of the first step's volume of parameters are architected to be experimented, following the gradient of better performance and complying with the DOE methodology. Thereby, the second step of the standard operation mode's DOE was planned to explore the volume of parameters for the region of high level of *timesleep*, high level of *interval* and high level of *frequency*, with respect to the first step, since these configurations have led to performances with the smaller ASSD values. The choice of the second step parameters' values was so defined that the new step explores the half of values of the first step's levels of parameters. For instance, in the first step, the factor *frequency* has low level of 5 and high level of 120. Since the best performance achieved was according to performances in high level of *frequency*, the new boundaries of this factor in the second step is consisted of exploring the half of values closer to 120. Thus, since the value 63 is roughly in the middle of the closed interval $[5, 120]$, the new values of this parameter for the second step will stand as, low level equal to 63, and high level equal to 120. This reasoning is applied for the choice of values for the other parameters, and consists the strategy of following the gradient of better performance.

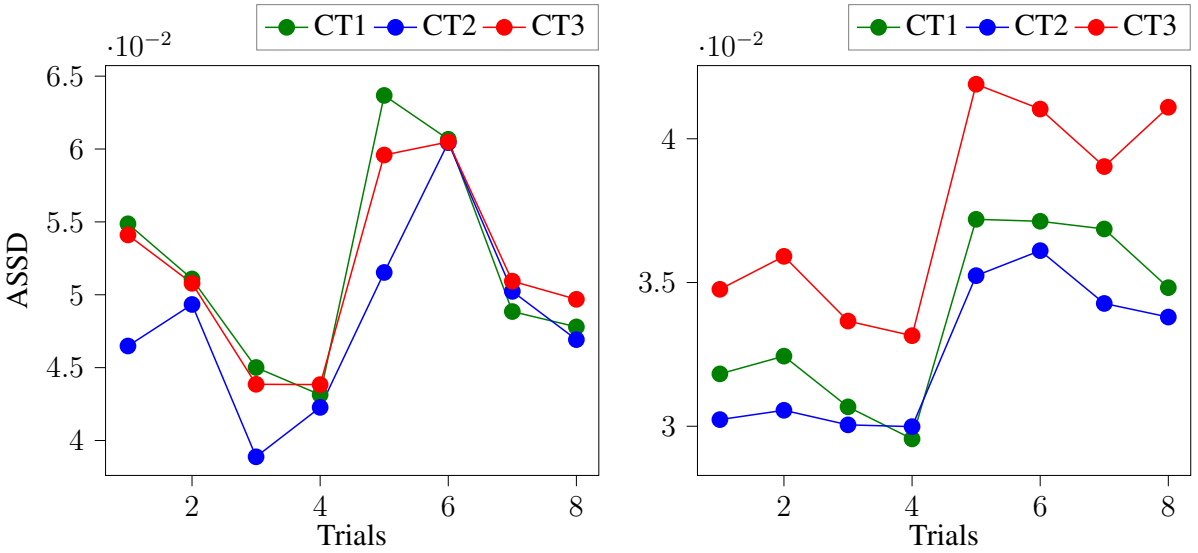
The second step of the standard operation mode's DOE is depicted as the first inner cube in Figure 34, for the sensing x-axis, and in Figure 35 for the sensing y-axis. An immediately general assessment of this step through the cube plots conveys the information that this volume of parameters leads to much more stable performances, for both the sensing axes. For the sensing x-axis, the ASSD values range from 0.043 to 0.061, whereas the range 0.031 to 0.038 stands for the sensing y-axis. In comparison with the first step, the scale of the ASSD values dropped drastically, and thus, better were the performances of the second step. An interesting observation is remarked regarding a comparison between the two sensing axis. In the first step, the sensing y-axis achieved ASSD values smaller than the sensing x-axis, for the same combination of parameters. This pattern remains in the responses of the second step. Furthermore, in the second step, the range of the ASSD values is smaller for the sensing y-axis, than for the sensing x-axis. These considerations seem to attest that the sensing y-axis achieves performances more stable and accurate than the sensing x-axis, for any set of parameters. Another interesting observation for the results of this step is that the gradient of better performance seems to point to low level of *frequency* and high level of *interval*. Differently from the first step, the varying of the parameter *timesleep* seems to lead in the second step to insignificant changes of performance.

Table 26 – Table presenting the mean and standard deviation value, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the second step belonged to the standard operation mode.

Trial number	**ASSD mean x-axis**	**Std x-axis**	**ASSD mean y-axis**	**Std y-axis**
1	0.0518	0.0046	0.0322	0.0022
2	0.0504	0.0009	0.0329	0.0027
3	0.0425	0.0032	0.0314	0.0019
4	0.0430	0.0007	0.0308	0.0019
5	0.0582	0.0061	0.0381	0.0034
6	0.0605	0.0001	0.0380	0.0025
7	0.0500	0.0010	0.0367	0.0023
8	0.0481	0.0014	0.0365	0.0039

Concerning the variability of the conjunct of trials executions' responses from the second step, the Figure 38 and Table 26 present the performance of each trial according to its respective conjunct of trials. Fig.38 left, presents the responses of the sensing x-axis, whereas Fig.38 regards the sensing y-axis. The trial's sequence number stands for the trial's design from Table 16. Table 26 conveys the mean and standard deviation values of the respective trials from each conjunct of trials belonging to the second step. Once more, one may remark that the scale of the respective trials' values remained the same, for all the trials. In fact, the variability of the trials, conveyed through the standard deviation value, is quite similar to the variability from the first step. However, since the responses from the sensing y-axis achieved values quite akin among themselves, the difference between two different trial numbers' responses, for this sensing axis,

Figure 38 – Graphics presenting the results of the ASSD metric for all the three conjuncts of trials performed in the second step. The number of trials, found in the abscissa axis, stands for the numeration of the trial's design from Table 16. The graphic from figure left, regards the response for the sensing x-axis, whereas the graphic in the right concerns the sensing y-axis. The abbreviation 'CT' refers to Conjunct of Trials.



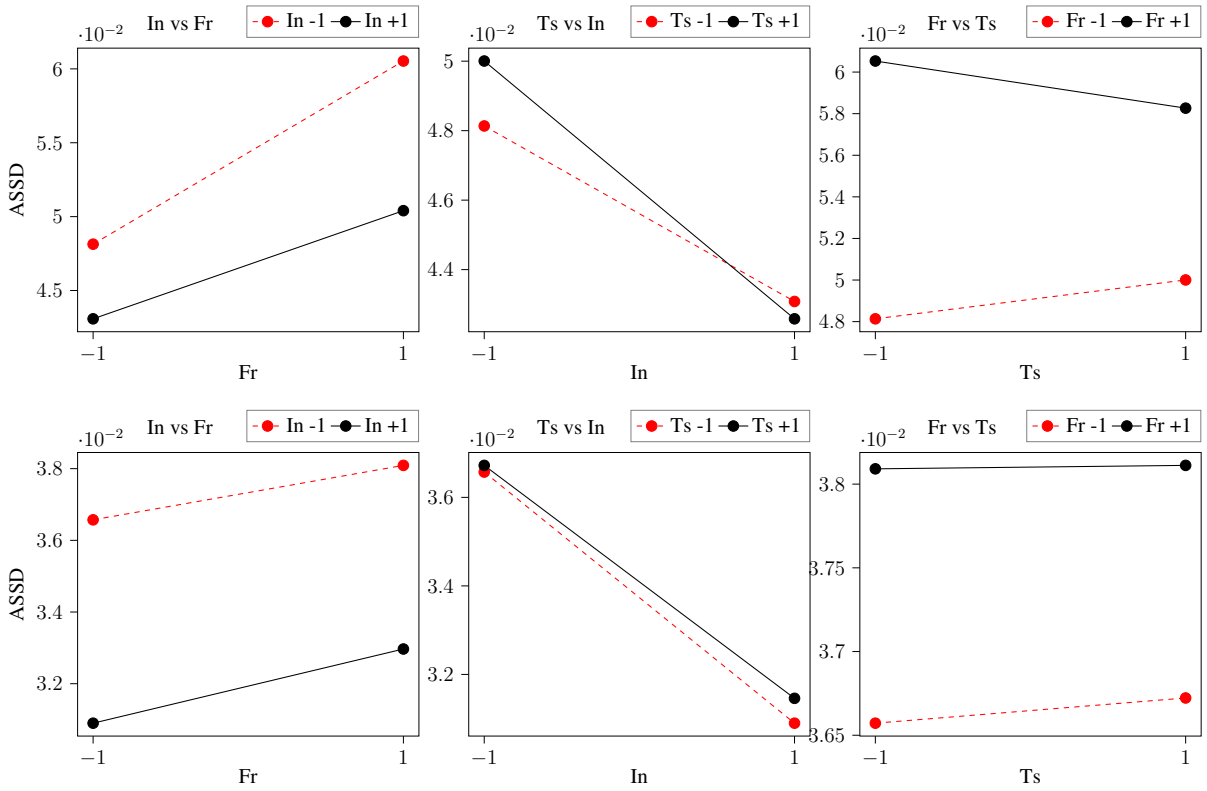
Source: Author.

edges the uncertainty of an individual trial number's response, assessed here through the standard deviation, and thus, the uniqueness of each trial number's performance in the sensing y-axis is compromised. This phenomenon is not observed in the responses concerning the sensing x-axis, though the uncertainty values from both the sensing axes are similar, since the difference between two different trial numbers' responses in the sensing x-axis is not close to the trials' respective uncertainty.

The Figure 39 presents the interactions plots for the second step. The first row of three plot's set stands for the sensing x-axis' responses, whereas the second row for the sensing y-axis' responses. Since the results concerning the sensing y-axis' responses point to unreliability regarding the performance's uniqueness of this sensing axis at the second step, the following analyses concerning the interactions of the sensing y-axis are assumed as a suggestion of the device's behaviour.

Evaluating the interactions plots from the first row of plots' set, one may remark that the degree of non-parallelism between all the pairs of lines is quite subtle, differently from the interactions plots of the first step, in Figure 37, where the degree of non-parallelism is pretty notable. The interactions responses of the sensing y-axis, in the second row of Fig.39, are found considerable akin to the interactions of the sensing x-axis, save for the fact that the degree of non-parallelism of the responses belonging to the sensing y-axis seems to be almost insignificant. The interactions plots of the second step convey the knowledge that the behaviour of the parameters' interactions changed from the first step to the second step, not just regarding

Figure 39 – Interactions plots of the second step of experiments from the standard operation mode. The first row of group of three plots stands for the ASSD values of the sensing x-axis' performance, whereas the second row stands for the sensing y-axis' performance. For each interaction plot, the remnant parameter is set to low level.



Source: Author.

the degree of non-parallelism, but also the direction of the gradient of better performance seems to had changed for the interaction between *interval* and *frequency*. For the first step, assessing the interaction between *interval* and *frequency*, for both the sensing axes, the smaller ASSD values are found when the factor *frequency* is set to high level. This pattern changed in the second step, where smaller ASSD values are found, for the interaction between *interval* and *frequency*, when the factor *frequency* is set to low level. These observations convey the fact that the interactions among the parameters, for a volume of parameters delimited by the frontiers of the first step (Table 15), are slightly non-linear. The behaviour of the parameters' interactions of the inclinometer change according to the inner volume of parameters analysed. Thereby, the following of better performance's gradient is inaccurate to find a sweet spot, due to the fact that the approach deployed to follow this gradient is based in to assess the half of parameters that had led to the better responses. But since the interactions are non-linear, this gradient could have changed its direction at any point before the half of parameters' volume that is being set to be followed. The DOE approach can not provide much more information to find a sweet spot in a process that has non-linear parameters' interactions. It is a limitation of the methodology.

Although the following of better performance's gradient is no more a reliable approach

to find a sweet spot of combination of parameters that leads to better performances, since the interactions among the inclinometer's parameters were attested no linear through the delimited volume of parameters of the second step, yet a third step of standard operation mode's DOE was performed, in order to remark whether the conclusions so far achieved would be observed again.

The third step of DOE belonging to the standard operation mode is depicted as the smallest cube plot in Figure 34, for the sensing x-axis, and Fig.35 for the sensing y-axis. The plots convey the knowledge that the responses of the third step were pretty akin to the value of the ASSD metric from the second step, though the values for the third step seem to be slightly smaller. The ASSD values concerning the entire volume of parameters from the third step range from 0.043 to 0.054, for the sensing x-axis, and from 0.031 to 0.035 for the sensing y-axis. The short range of values for both the sensing axes suggests the hypothesis that the parameters' volume of the third step either reaches a spot of stagnating performance, or the levels of the factors are not spaced enough to unleash a notorious change in the measurement's performance. There is also the possibility that both the hypothesis are right.

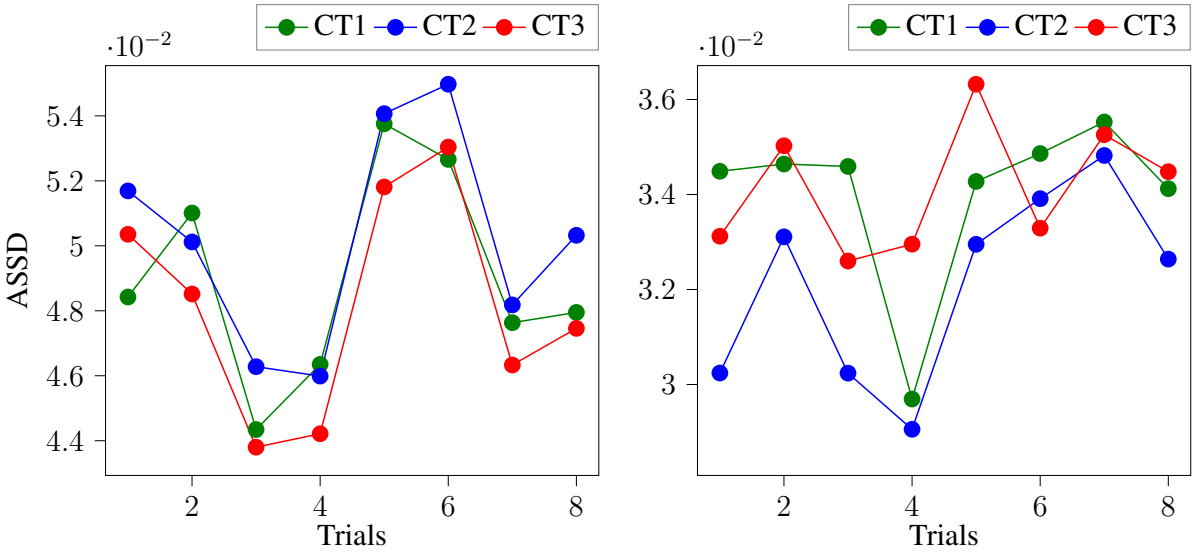
Table 27 – Table presenting the mean and standard deviation value, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the third step belonged to the standard operation mode.

Trial number	**ASSD mean x-axis**	**Std x-axis**	**ASSD mean y-axis**	**Std y-axis**
1	0.0501	0.0016	0.0326	0.0021
2	0.0498	0.0012	0.0342	0.0010
3	0.0448	0.0013	0.0324	0.0021
4	0.0455	0.0011	0.0305	0.0020
5	0.0532	0.0012	0.0345	0.0016
6	0.0535	0.0012	0.0340	0.0007
7	0.0473	0.0009	0.0352	0.0003
8	0.0485	0.0015	0.0337	0.0009

The Figure 40 and Table 27 convey the variability's performance of the conjunct of trials of the third step from the standard operation mode's DOE. Fig.40 left presents the response for the sensing x-axis, whereas Fig.40 right the response of the sensing y-axis. The standard deviation of the trials' values, regarding conjunct of trials among themselves, presented in Table 27, brings again the observation that, since the ASSD values from the third step are within a short range, for both the sensing axes, the uniqueness of each parameters combination's performance is compromised, because the range among the responses edges the uncertainty of the trials' responses, assessed here through the standard deviation.

To conclude the third step's evaluation, Figure 41 presents its interactions plot. The first row of plot's group stands for the results of the sensing x-axis, whereas the second row stands for the sensing y-axis. An immediate observation must be remarked before the analysis of the graphics. Since the uniqueness of performance is compromised for some set of parameters, for

Figure 40 – Graphics presenting the results of the ASSD metric for all the three conjuncts of trials performed in the third step. The number of trials, found in the abscissa axis stands for the numeration of the trial's design from Table 16. The graphic from figure left, regards the response for the sensing x-axis, whereas the graphic in the right concerns the sensing y-axis. The abbreviation 'CT' refers to Conjunct of Trials.

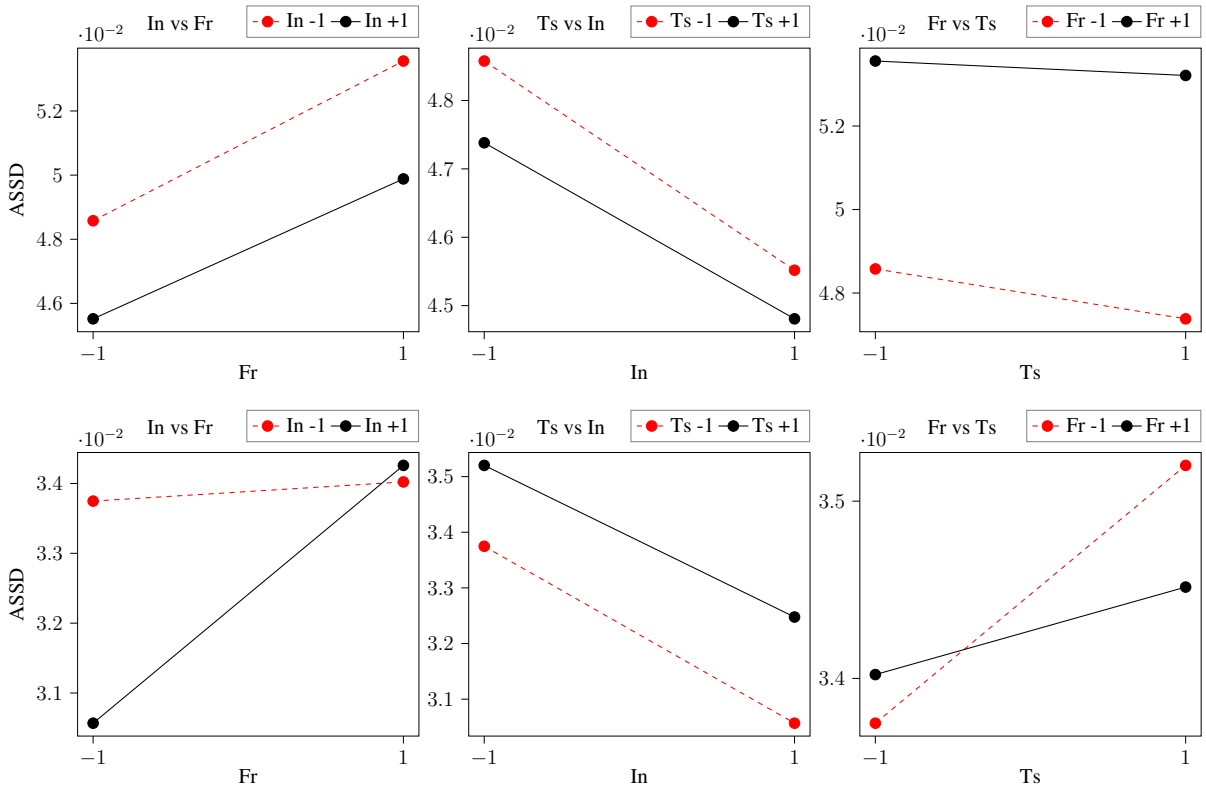


Source: Author.

both the sensing axes, no conclusions can be asserted from these interactions plots. However, is interesting to observe that, the interactions of the first row of plots suggests that the interactions among the parameters appear to converge to the state of no interactions, since the interaction lines assume a parallelism profile. Once that the responses of the sensing y-axis had smaller range of values than the sensing x-axis, its interactions plots appear to convey that there is interaction among the parameters, save for the interaction between *timesleep* and *interval*, which remained with the status of no interaction. Nevertheless, these analyses are not reliable. It will be assumed that the interactions plots of Figure 41 just suggests that the interactions among the parameters from the third step converge to the status of no interaction.

The three steps of the standard operation mode's DOE provide a characterisation of the operation mode, assessing its gradient of better performance for each conjunct of trials parameters' list. By the means of the cube plots here provided, one may have an intuition of which combination of parameters is better for a given application. The interactions plots of the three steps suggest the knowledge that the parameters interaction's behaviour changes for each volume of parameters. This fact leads to the observation that, finding a sweet spot of parameters' settings that lead to an optimum performance, assessed through the ASSD metric, is not a trivial endeavour. The variability among the respective trials numbers from each conjunct of trials' execution is proved small for the standard operation mode, but not regardless. Indeed, it became evinced that the degree of variability of a trial's ASSD number depends on the trial's parameters settings, since that high ASSD values achieved high standard deviation values with respect to

Figure 41 – Interactions plots of the third step of experiments from the standard operation mode. The first row of the group of three plots stands for the ASSD values of the sensing x-axis' performance, whereas the second row stands for the sensing y-axis' performance. For each interaction plot, the remnant parameter is set to low level.



Source: Author.

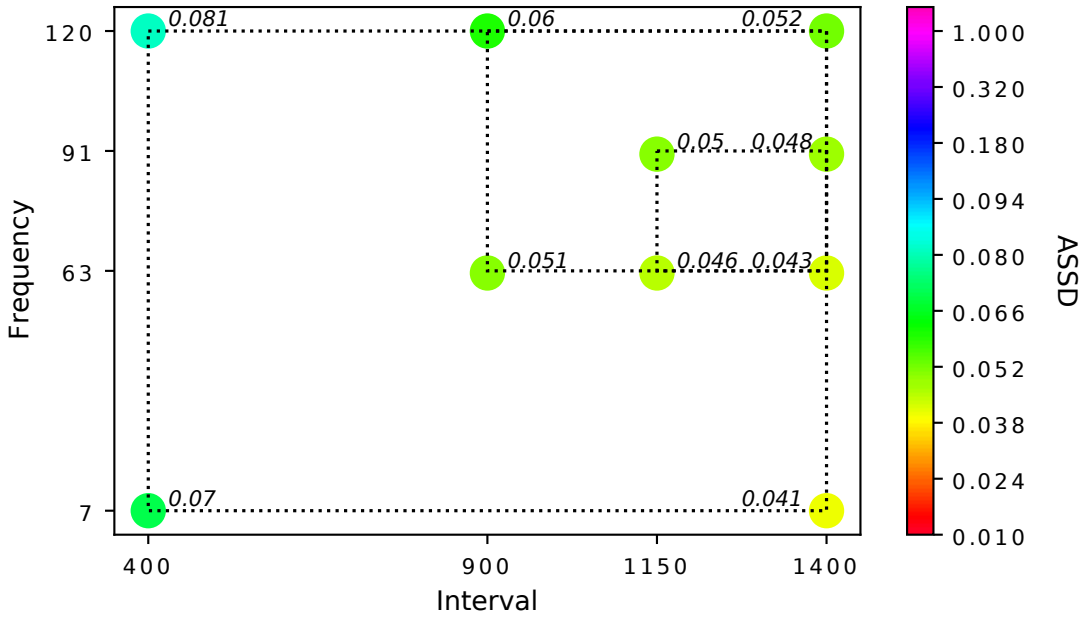
the same trials numbers belonging to other conjunct of trials' execution, and low ASSD values achieved low standard deviation values. The ASSD metric and the systematic methodology could manage to satisfactory remark the patterns of each trial, so that there is no need to carry out several runs of the same trial in order to remark an average of the ASSD values to state the pattern of a given combination of parameters, once that, although with variability, the ASSD values of each trial number kept the same pattern. The conclusions here presented are further employed to formulate the design of the machine learning approach for the standard operation mode, which aims to characterise an entire delimited volume of parameters.

B.3.2 Repeated operation mode DOE

The experiments concerning the repeated operation mode were structured as depicted in Figure 33. As in the standard operation mode's DOE, the experiments for the repeated operation mode also consist of steps, where the first step regards an outer volume of parameters, depicted in Table 19, the second step stands for combinations of parameters inner the volume of the first step, approached in Table 21, and the third step provides experiments with settings of parameters inner

the second step's volume, as conveys Table 22. The order of step's execution was sequential, from step one to step three.

Figure 42 – Square plot of the three steps of experiments from the repeated operation mode, representing the response of the ASSD metric for the sensing x-axis. The most outer delimited square stands for the first step, the middle inner square for the second step, and the smallest and innermost square for the third step. One may remark that, the closer to zero is the ASSD metric value, the more stable and smaller is the noise component of the response whose a parameters combination leads.

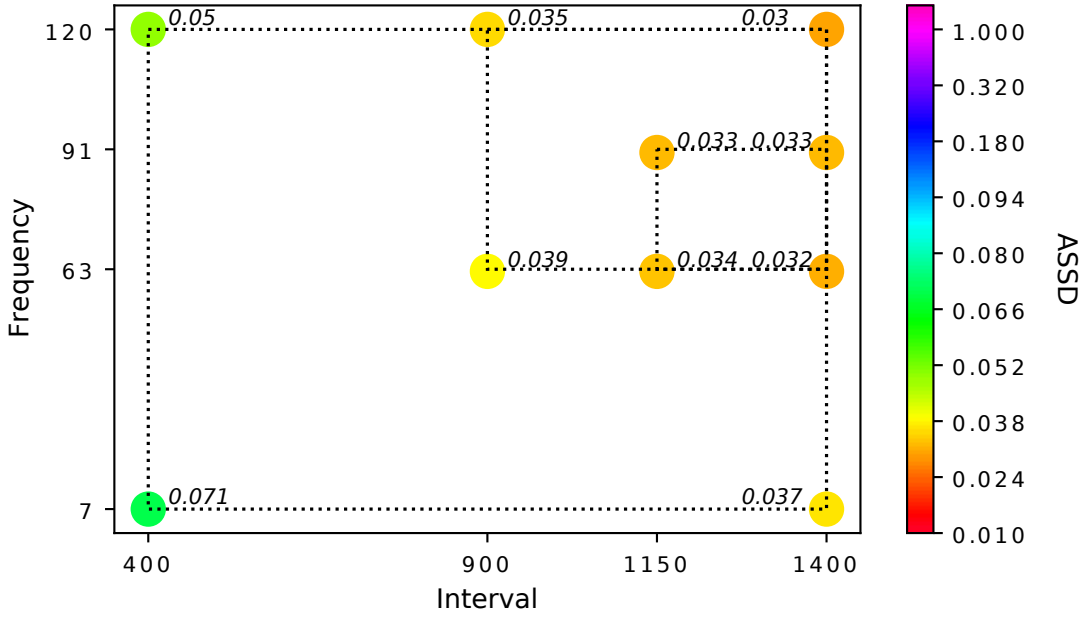


Source: Author.

Since the standard mode's DOE has three controllable variables, where each variable varies between two levels, it is said that its design is composed of a 3^2 architecture. The repeated operation mode's DOE is driven by two controllable variables, at two levels. Thus, this mode of operation is designed as a 2^2 architecture. Considering that it is employed the strategy of full factorial design, a 2^2 structure can be conveyed through a square plot. In order to easier assess the repeated operation mode's performance, the figures 42 and 43 present the square plots of its experiments. Figure 42 portrays the responses of the ASSD metric for the sensing x-axis, whereas Figure 43 stands for the sensing y-axis. Each coloured point in the square plots represents the average of the three trials' ASSD values, which refer for the performance of the three times executed conjunct of trials belonging to the step in analyses. As occurs for the cube plots in the analysis of the standard operation mode, since that in the square plots, for each inner volume of parameters, one set of parameters is shared by two square plots, whenever this graphic particularity occurs, the response of the most outer volume of parameters is displayed. The real value of the responses that were not exhibited can be found in the tables 29 and 30.

Evaluating the most outer square plot regarding the sensing x-axis, which stands for the first step, an immediate contrast with the standard operation mode first step's DOE is remarked.

Figure 43 – Square plot of the three steps of experiments from the repeated operation mode, representing the response of the ASSD metric for the sensing y-axis. The most outer delimited square stands for the first step, the middle inner square for the second step, and the smallest and innermost square for the third step. One may remark that, the closer to zero is the ASSD metric value, the more stable and smaller is the noisy component of the response whose a combination of parameters leads.



Source: Author.

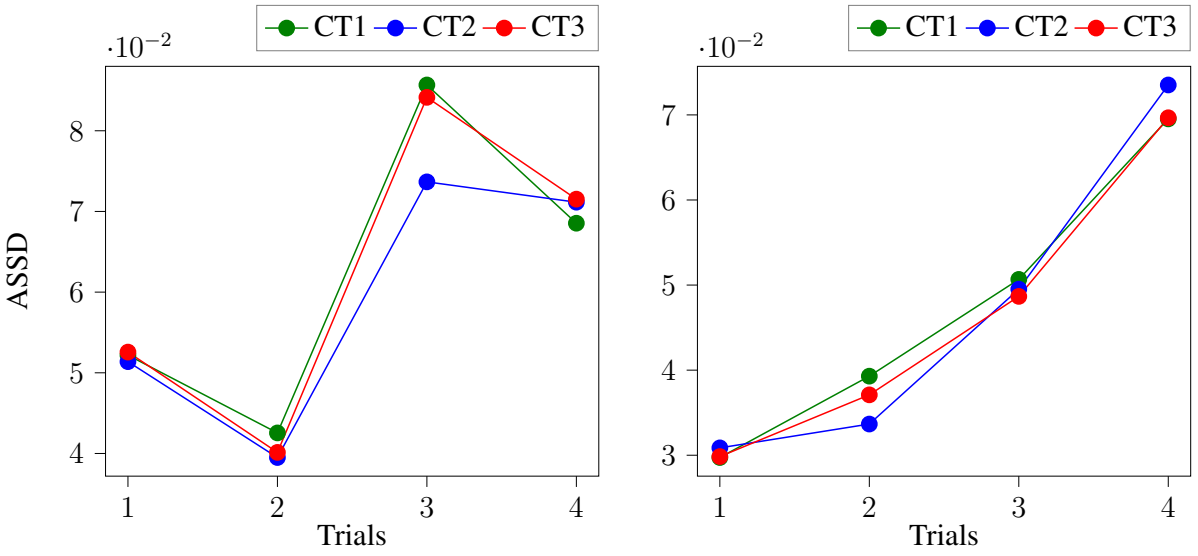
The ASSD values for the repeated operation mode's first step are drastically smaller than the same step for the standard mode. In fact, the scale of the values from the repeated first step competes with the performances regarding the standard second step. In Figure 42 one may also observe that the parameter *interval* seems to have a considerable influence on the performance's behaviour. Better performances for the sensing x-axis of this first step are found when the factor *interval* is set to high level. The parameter *frequency* seems to lead to better performances when is set to low level. However, an interesting pondering can be remarked for the first step of the sensing y-axis' responses (Fig.43), when compared with the same responses of the sensing x-axis. For the sensing y-axis, the parameter *interval* also has the majority influence on the performance's driving, leading to better responses when set to high level. However, for the parameter *frequency*, differently from the responses of the sensing x-axis, better performances were achieved when the factor *frequency* was set to high level. These results suggest that the factor *frequency* has an inverse-proportional relationship concerning each sensing axis. This condition leads to a laborious effort when one is interested in to find a sweet spot of combination of parameters that leads to the best performances, since one would need to set a trade-off between the performance of both the sensing axis.

The Figure 44 and Table 28 present the performance of the variability among the

Table 28 – Table presenting the mean and standard deviation value, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the first step belonged to the repeated operation mode.

Trial number	ASSD mean x-axis	Std x-axis	ASSD mean y-axis	Std y-axis
1	0.0520	0.0006	0.0301	0.0006
2	0.0407	0.0016	0.0366	0.0028
3	0.0811	0.0065	0.0496	0.0010
4	0.0704	0.0016	0.0709	0.0022

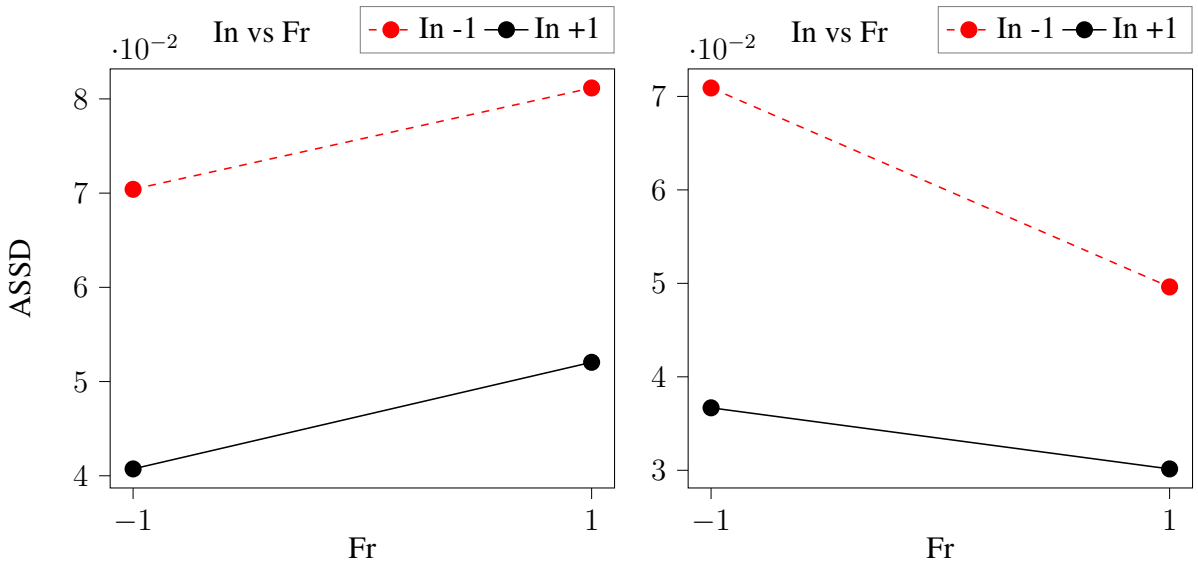
Figure 44 – Graphics presenting the results of the ASSD metric for all the three conjuncts of trials performed in the first step. The number of trials, found in the abscissa axis, stands for the numeration of the trial's design from Table 20. The graphic from Figure left, regards the response for the sensing x-axis, whereas the graphic in the right concerns the sensing y-axis. The abbreviation 'CT' refers to Conjunct of Trials.



Source: Author.

respective trials from the three conjunct of trials' execution that consists the first step. Table 28 informs the mean value of a trial's number, regarding the design from Table 20, and its associated standard deviation. Figure 44 conveys visually the performance of each conjunct of trials. Fig.44 left present the response for the sensing x-axis, whereas fig.44 right stands for the sensing y-axis. One may observe from the graphics that all the conjunct of trials had performances pretty akin among themselves. In fact, Table 28 conveys numerically, through the standard deviation, the degree of similarity regarding each number of trial. Since all the standard deviations values are not found high, one may conclude that the systematic methodology could remark the performance's patterns whose each combination of parameters had led with reasonable low variability.

Figure 45 – Interactions plots of the first step of experiments from the repeated operation mode. The Fig. left stands for the ASSD values of the sensing x-axis' performance, whereas the Fig. right stands for the sensing y-axis' performance.



Source: Author.

The Figure 45 presents the interactions plots from the first step of the repeated operation mode's DOE. Fig.45 left depicts the interaction between the factors *interval* and *frequency*, for the sensing x-axis, whereas Fig.45 right stands for the sensing y-axis. For both the sensing axis, the interaction between *interval* and *frequency* seems to be stated as non-interaction, since the interaction lines may be considered parallel, despite the subtle non-parallelism found in the lines belonging to the sensing y-axis' response. Figure 45 also provides a visualization of the gradient of better performance for the repeated mode's first step. One may note that the ASSD's smaller values are found when the interaction lines point to low level of *frequency*, for the sensing x-axis, and high level of *frequency* for the sensing y-axis. Since no interaction is considered for the repeated operation mode's factors, it suggests that the endeavour to find a sweet spot of better settings of parameters that lead to better performances is contemplated just for the finding of a trade-off of settings, so that the frequency's inverse-proportional relationship between the two sensing axis is compensated.

In order to follow the DOE methodology to find a sweet spot of combinations of parameters that leads to better possible performances, the experiments with the repeated operation mode were explored into steps, following the same standard described for the standard operation mode. The choice of parameters that consists the second step's DOE, presented in Table 21, was so defined that it complies with the following of better performance's gradient of the parameter interval. Since the same levels of the parameter *frequency* led to different behaviours for the two sensing axis, the frequency's gradient of better performance is not the same for the two sensing axis. Thus, it was chosen, without a particular reason, to follow the frequency's gradient of better

performance of the sensing y-axis. Therefore, the volume of parameters that consists the second step is composed by the exploring of the half of values closer to the high level of *interval* and *frequency*, regarding the first step.

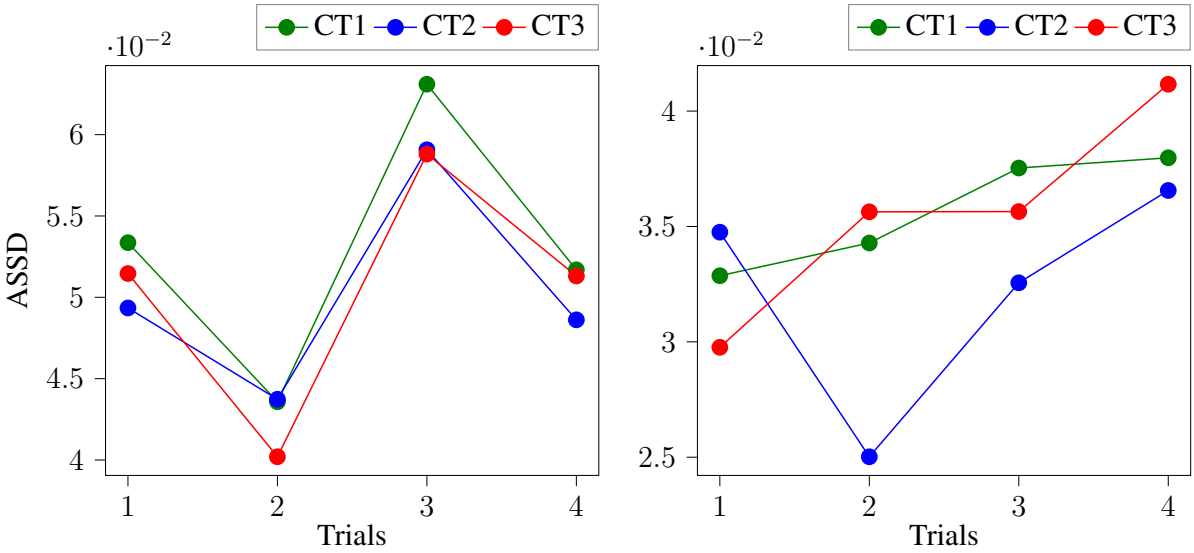
The second step of repeated operation mode's experiments can be visually assessed through the first inner square from figure 42 and 43. One may remark that, comparing the second step with the first step, for both the sensing axis, the performance's patterns seem to have been preserved. The sensing x-axis' responses had better performances when *interval* is set to high level, and *frequency* to low level. For the sensing y-axis, the better responses are found also when *interval* is set to high level, but, differently from the sensing x-axis, and keeping the inverse-proportional frequency's relationship between the sensing axes, better performances were achieved when the factor *frequency* was set to high level. All the performances belonging to the second step accomplished smaller ASSD values than the experiments belonging to the first step. One may also note that, the smaller ASSD values are found, for the first and second step, in the responses from the sensor y-axis. This pattern was also remarked for the experiments regarding the standard operation mode. Such facts comply with the suggestion that the sensing x-axis is constrained to achieve worse responses than the sensing y-axis.

Table 29 – Table presenting the mean and standard deviation value, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the second step belonged to the repeated operation mode.

Trial number	**ASSD mean x-axis**	**Std x-axis**	**ASSD mean y-axis**	**Std y-axis**
1	0.0513	0.0020	0.0324	0.0025
2	0.0425	0.0019	0.0316	0.0057
3	0.0603	0.0024	0.0352	0.0025
4	0.0505	0.0016	0.0385	0.0023

The Figure 46 and Table 29 convey the variability of response regarding each individual trial number with its respective trials from the other conjuncts of trials belonging to the second step of the repeated operation mode. The trials' number and sequence stand for the DOE design depicted in Table 20. In Figure 46 left, which regards the sensing x-axis, one may observe that each trial number was distinctly remarked by the ASSD metric and the systematic methodology, with a reasonable variability. Differently for the responses from the sensing y-axis, Fig.46 right, where the response of each trial number is pretty akin to the response of other trial numbers. In fact, the range among the different trials number's responses edges the uncertainty of an individual trial number's response, assessed here through the standard deviation, and hence, the uniqueness of each trial number's performance is compromised for the sensing y-axis. This phenomenon is also observed in the second and third step of the standard operation mode, and leads to the conclusion that the repeated operation mode's second step volume of parameters achieves, for the sensing y-axis, either a spot of stagnating performance, or the levels of the factors are not spaced enough to provoke a notorious change in the measurement's performance.

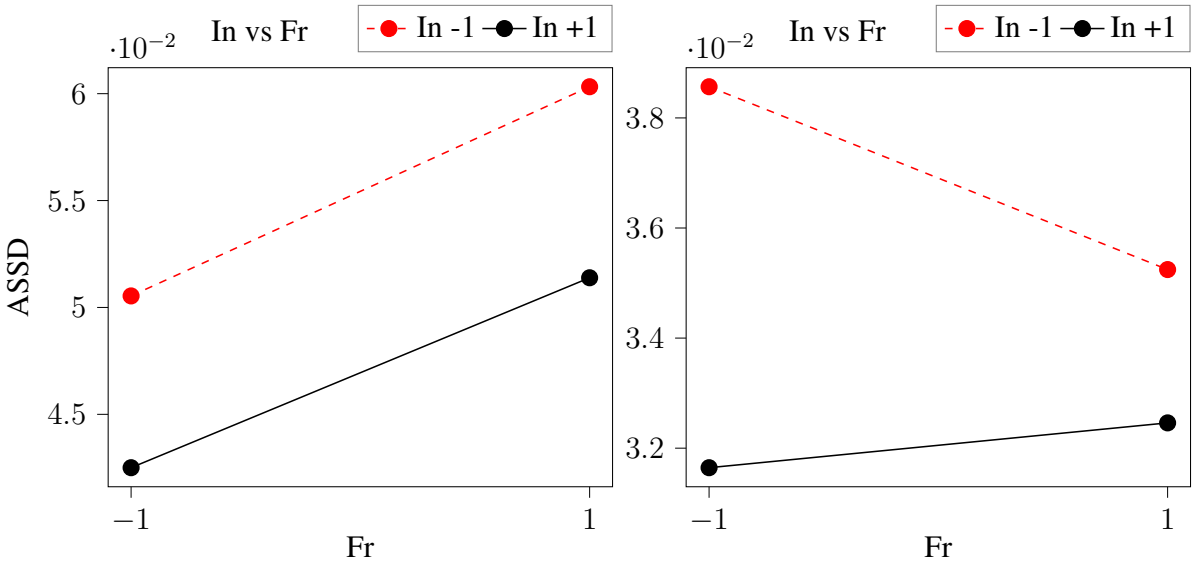
Figure 46 – Graphics presenting the results of the ASSD metric for all the three conjuncts of trials performed in the second step. The number of trials, found in the abscissa axis, stands for the numeration of the trial's design from Table 20. The graphic from Figure left, regards the response for the sensing x-axis, whereas the graphic in the right concerns the sensing y-axis. The abbreviation 'CT' refers to Conjunct of Trials.



Source: Author.

And as previously noted, there is also the possibility that both the hypothesis are right.

Figure 47 – Interactions plots of the second step of experiments from the repeated operation mode. The Fig. left stands for the ASSD values of the sensing x-axis' performance, whereas the Fig. right stands for the sensing y-axis' performance.



Source: Author.

The Figure 47 portrays the interaction plot regarding the repeated operation mode's second step. Since the results concerning the variability of the responses from the sensing y-axis

conveyed the knowledge that the performance's uniqueness of this sensing axis is compromised for the second step, the interactions plots' analyses remarked for the sensing y-axis are not reliable. However, the analyses concerning the sensing y-axis are further presented as a suggestion of the device's behaviour of the volume of parameters from the second step.

Comparing the interactions from the second step with those from the first step, one may remark that the same pattern is maintained, save for the interactions of the sensing y-axis, which in the first step were considered with the state of non-interaction, and here, in the second step, can not be as well stated, since the degree of non-parallelism between the interaction lines has increased considerably. Thus, it is said that the factors of the sensing y-axis have interactions. The change of interactions' behaviour from one step to another is also observed in the first and second step from the standard mode. The occurrence of this phenomenon points out to the conclusion that the interactions of the inclinometer parameters have a non-linear relationship, so that the following of the better gradient of performance aiming the finding of a sweet spot of parameters is impaired.

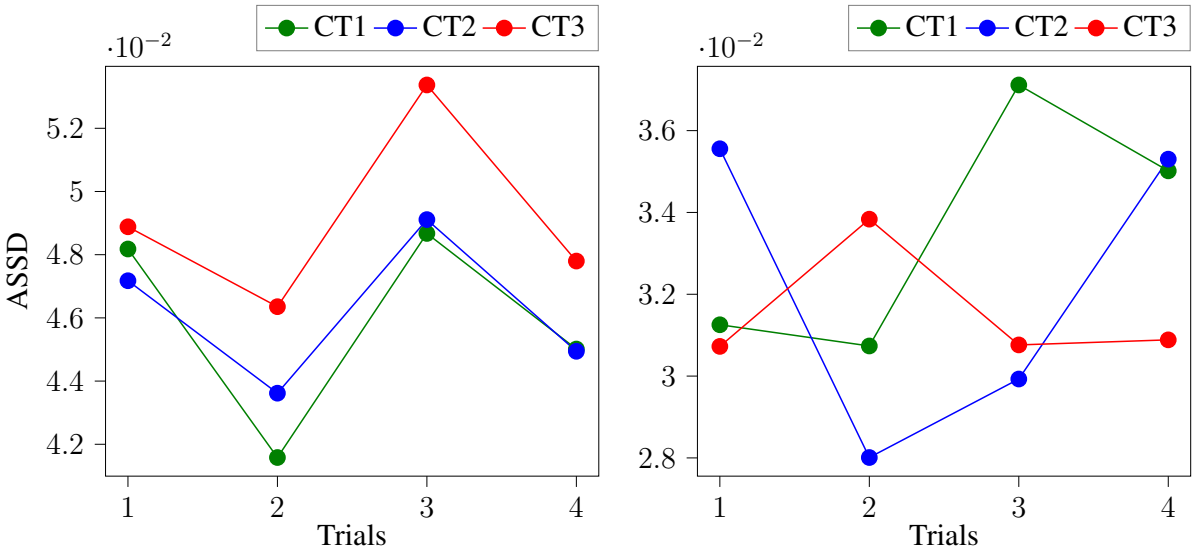
In order to remark whether the patterns so far observed would remain for other volume of parameters, the third step of the repeated operation mode's DOE was so designed that the better performance's gradient is followed. The third step volume of parameters was defined through setting new boundaries of parameters with the half of values closer to the parameters that had led to better performances of the sensing x-axis at the second step. Thereby, the new boundaries follow the high level of *interval* and low level of *frequency* from the second step. The third step of the repeated operation mode's DOE is depicted as the innermost square in Figure 42, for the sensing x-axis' responses, and Figure 43 for the sensing y-axis' responses. Comparing the ASSD values of the third step with the second step, one may remark an improvement in the performances belonging to the sensing x-axis. For the sensing y-axis, the responses achieved roughly the exact same values as the previous step, but slightly improved. This observation complies with the pattern of all the steps so far assessed, regarding the standard and repeated operation mode, where the sensing y-axis achieved notably better responses than the sensing x-axis. Better performances of the sensing x-axis are found when *interval* is settled to high level, and *frequency* to low level. Such exact pattern had been observed in the first and second step for the responses of this sensing axis. Since the ASSD values belonging to the sensing y-axis are approximately the same in this step, it is not possible to reliably identify the gradient of better performance for this step.

The Figure 48 and Table 30 convey the performance's variability of each trial number regarding the three conjunct of trials' executions belonging to the experiments of the third step. Once again, the systematic methodology could mark the behaviour's patterns of the sensing x-axis' responses, Fig.48 left, with reasonable accuracy. Differently for the sensing y-axis' responses, Fig.48 right, where the systematic methodology could not assign a unique pattern for each trial number. The volume of parameters chosen for the third step seems to have led the sensing y-axis' performance to a stagnated spot of achievement. One may observe that the

Table 30 – Table presenting the mean and standard deviation value, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the third step belonged to the repeated operation mode.

Trial number	ASSD mean x-axis	Std x-axis	ASSD mean y-axis	Std y-axis
1	0.0480	0.0008	0.0325	0.0026
2	0.0438	0.0023	0.0308	0.0029
3	0.0503	0.0025	0.0326	0.0039
4	0.0459	0.0016	0.0337	0.0024

Figure 48 – Graphics presenting the results of the ASSD metric for all the three conjuncts of trials performed in the third step. The number of trials, found in the abscissa axis, stands for the numeration of the trial's design from Table 20. The graphic from Figure left, regards the response of the sensing x-axis, whereas the graphic in the right concerns the sensing y-axis. The abbreviation 'CT' refers to Conjunct of Trials.

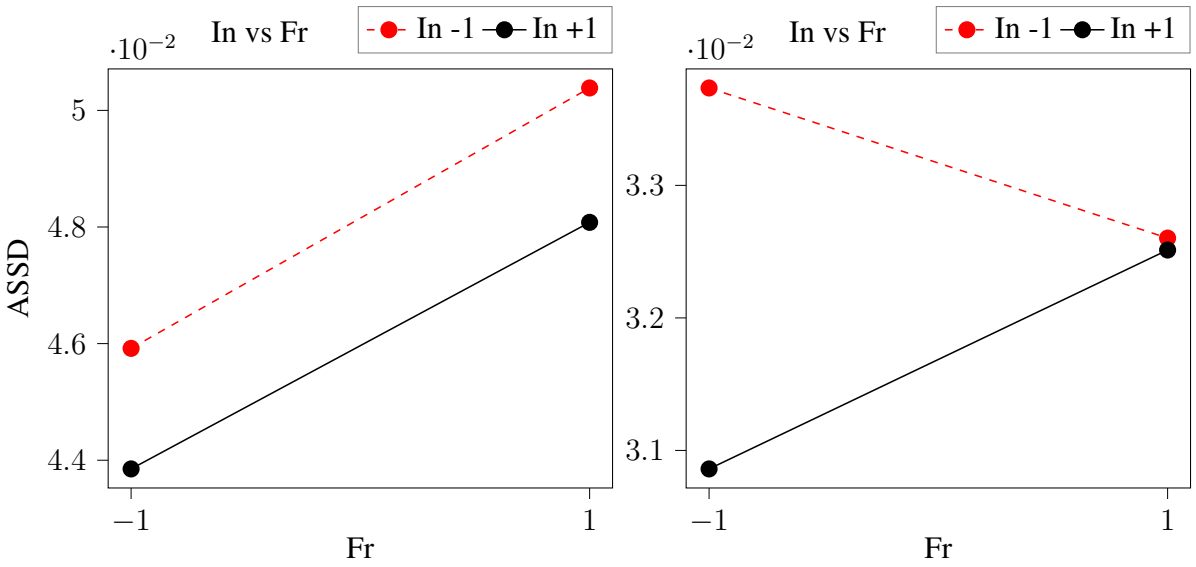


Source: Author.

sensing y-axis' mean ASSD values, Table 30, of all the four trials are found with roughly the same value, with an uncertainty of measurement assessed through the standard deviation, not neglectable. Thereby, the performance's uniqueness of the sensing y-axis is compromised in the third step, since the range among the different trials number's responses edges the uncertainty of a individual trial number's response. Such phenomenon is not observed for the sensing x-axis' responses in this step, which reinforce the hypothesis that the third's step settings lead to a spot of stagnating performance of the sensing y-axis, and undermines the risen hypothesis that the levels of the factors are not spaced enough to provoke a notorious change in the measurement's performance.

The Figure 49 presents the interactions plot of the repeated operation mode's third step

Figure 49 – Interactions plots of the third step of experiments from the repeated operation mode. The Fig. left stands for the ASSD values of the sensing x-axis' performance, whereas the Fig. right stands for the sensing y-axis' performance.



Source: Author.

DOE. Fig.49 left stands for the interactions regarding the sensing x-axis' responses, whereas Fig.49 right for the sensing y-axis' responses. Once again, the interactions plot concerning the sensing y-axis is not reliable, since that the variability analysis conveyed the knowledge that the performance's uniqueness of this sensing axis is compromised for this step. Nevertheless, the following assessment concerning the sensing y-axis is presented as a suggestion of the device's behaviour for the volume of parameters from the third step.

The interaction between *interval* and *frequency* of the sensing x-axis, Fig.49 left, maintained its pattern of non-interactions, observed in the first and second step, once that the interaction lines remained parallel. In another hand, the responses for the sensing y-axis changed their pattern once again, with reference to the first and second step. The sensing y-axis interactions, Fig.49 right, seem to have been reinforced, since the degree of non-parallelism between the interaction lines increased.

The repeated operation mode's three steps DOE so far presented convey a characterisation of the inclinometer's performance for the operation mode under analyses. The systematic methodology could remark the device's behaviour with reasonable low variability, so that one is enabled to have an intuition of which combination of parameters is better for an application. The square plots portray a map of the sensor's performance according to the parameters set. The interaction plots convey the knowledge that the device maintains the same parameter's interaction behaviour during the three steps for the sensing x-axis. Since the sensing y-axis responses seemed to have achieved a stagnating spot of performance in the second and third step, where the uncertainties of the ASSD measurements edge the difference between

two different trial numbers' responses, no conclusions can be reliably taken concerning the sensing y-axis' interactions in the second and third step. The variability among the respective trials numbers from each conjunct of trials is not high for the repeated operation mode. Since the systematic methodology could manage to remark the patterns of each trial number for the repeated operation mode, there is no need to perform several runs of the same trial in order to remark an average of the ASSD values to state the performance's pattern for a given set of parameters. The conclusions here presented are further employed to substantiate the design of the machine learning approach for the repeated operation mode, which aims to characterise an entire delimited volume of parameters.

B.3.3 Continuous operation mode DOE

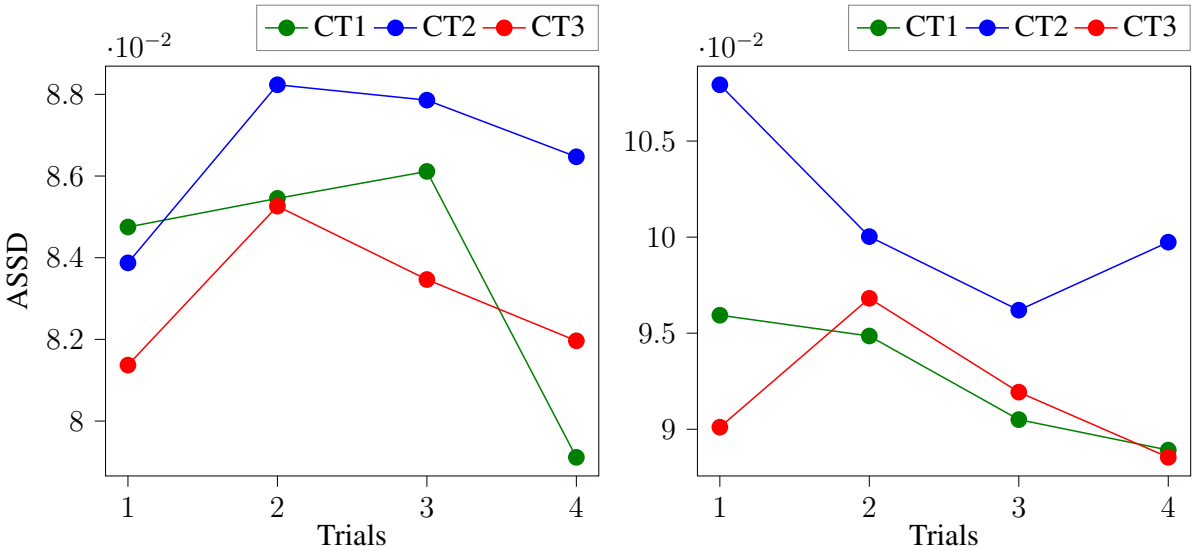
The continuous operation mode's DOE consists of just varying the parameter *frequency*. Since just one factor is varying, no full factorial design is applied, and thus, no cube/square plots or interactions plots are possible to be employed. Hence, the continuous operation mode is assessed just by the means of sequential graphics portraying the performance of each trial number regarding its respective conjunct of trials, Figure 50, and by the contemplation of the mean ASSD values from the respective trial numbers belonging to the three conjunct of trials, along with their standard deviation, presented in the Table 31. Due to the simplicity of the continuous' DOE model, there is no need to perform an exploration of the device's behaviour for inner parameters volume, once that the first and unique step of its evaluation can fully characterise the device's performance for this operation mode, because the step resolution configurable by the parameter *frequency* is 1 Hz, and the maximum stable value that this parameter can reach in the continuous mode is 4 Hz (Subsection 3.1.1). Thus, the delimited volume of parameters from Table 23 fully characterises the performance of this mode of operation.

Table 31 – Table presenting the mean and standard deviation value, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the first , and unique, step of experiments belonged to the continuous operation mode.

Trial number	ASSD mean x-axis	Std x-axis	ASSD mean y-axis	Std y-axis
1	0.0833	0.0017	0.0979	0.0090
2	0.0863	0.0016	0.0972	0.0026
3	0.0858	0.0022	0.0928	0.0029
4	0.0825	0.0037	0.0923	0.0063

The Figure 50 conveys the variability and scale of the trial numbers' ASSD values. Through the figure, one may remark that the variability among the respective trial numbers is quite similar to the variability observed so far in the other operation modes' DOE, however the standard deviations of the respective trial numbers, presented in Table 31 with the abbreviation 'Std', appear to be slightly higher than for the performance of the other modes of operation. The scale of the ASSD values, for both the sensing axis, is considerably high in comparison with

Figure 50 – Graphics presenting the results of the ASSD metric for all the three conjuncts of trials performed in the firs, and unique, step. The number of trials, found in the abscissa axis, stands for the numeration of the trial's design from Table 24. The graphic from figure left, regards the response of the sensing x-axis, whereas the graphic in the right concerns the sensing y-axis'. The abbreviation 'CT' refers to Conjunct of Trials.



Source: Author.

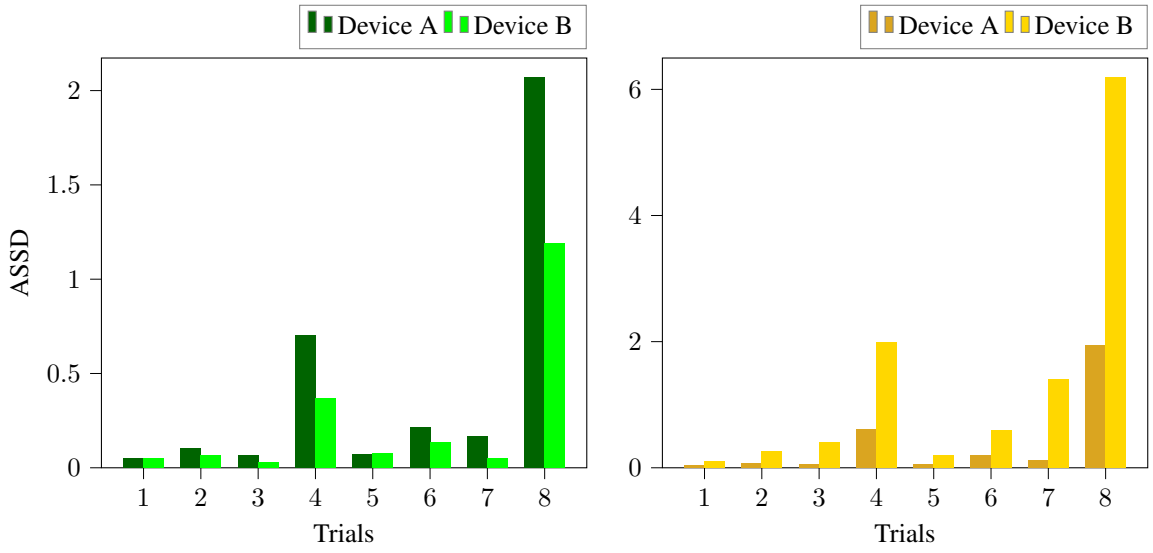
the responses of the repeated operation mode. Concerning the responses of each sensing axis, differently from the other modes of operation, the sensing x-axis presents better performances, for every trial number, than the sensing y-axis. Also, the Table 31 conveys the knowledge that all the trial numbers achieved roughly the same value, for the sensing x-axis, and quite similar values, for the sensing y-axis. The observations regarding the continuous operation mode lead to the conclusion that, regardless of the level of the factor *frequency*, the continuous operation mode presents approximately the same performance.

B.4 COMPARISON BETWEEN TWO INCLINOMETER DEVICES

In Subsection B.1 was briefly commented the fact that the strategy of replication could not have been applied in the inclinometer's DOE accomplished, due to the fact that two inclinometer devices had presented different performances for the same volume of parameters. In this section, the response of the two devices, which the author had in the time of the experiments, are evaluated and compared for the same combination of parameters.

It is defined as Device A, the inclinometer device that was employed in this work. The DOE analysis and the machine learning technique were deployed to the responses concerning the Device A. It is defined as Device B, another inclinometer device, that was just used in this work in order to compare the responses with the device A. In Figure 51 and Table 32, is presented a

Figure 51 – Figure presenting the performances of two inclinometer devices. The experiments were carried out through the exact same systematic performance from Section 3.4, so that both are analysed under the same conditions. Both the devices perform measurements according to the volume of parameters regarded by the Table 15, which concerns the first step of the standard operation mode's DOE. The device A refers to the device that was employed in this work. All the experiments of the DOE and machine learning approach were carried out by the device A. The device B was just employed in this work in order to compare the responses with the device A.



Source: Author.

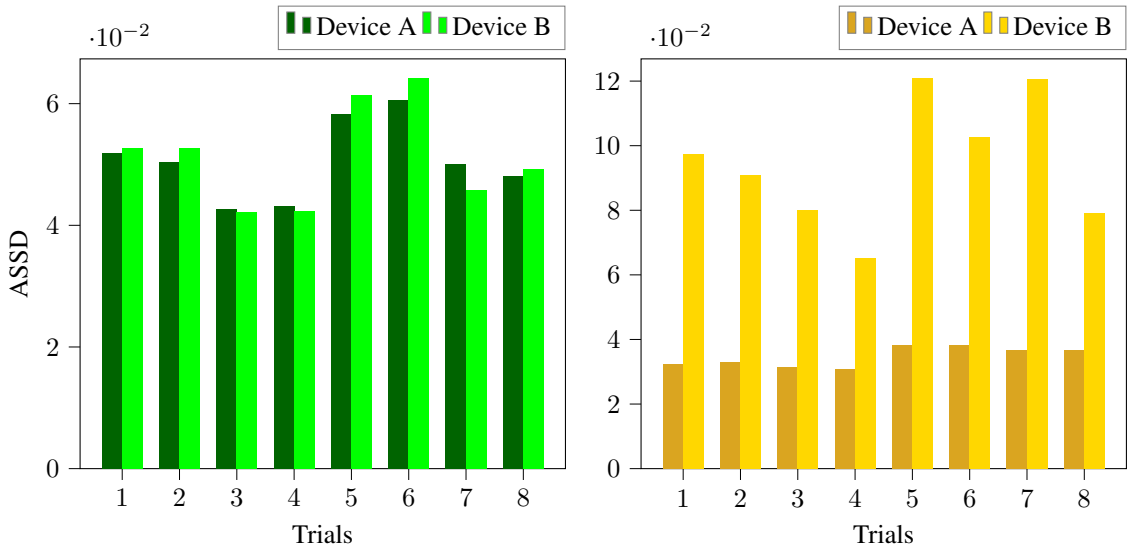
comparison of the responses from both the devices, for the combination of parameters regarding the first step of the standard operation mode (Table 15). Figure 51 left, presents responses concerning the sensing x-axis, whereas Figure 51 right stands for the sensing y-axis. The data regarding the Device A is the same depicted and assessed in the analyses of the standard operation mode in Subsubsection B.3.1. The Device's B responses were acquired through the exact same systematic methodology (Section 3.4), applying the principles of randomization and blocking (Subsection B.1), so that both the device's responses were taken under the same conditions, procedure and data processing. The ASSD trial values presented in figure 51 and Table 32 are the average of the ASSD response from each respective trial number belonged to the three executions of conjunct of trials performed in the first step, exactly as also conveyed for the responses of Subsection B.3.

Assessing Figure 51, one may clearly remark that the devices A and B achieve different responses for the same trial number, in both the sensing axis. In fact, the responses for the sensing x-axis, Figure 51 left, appears to follow the same gradient of better performance for both the devices, but the scale of the responses are mostly different. In contrast, the responses for the sensing y-axis, Figure 51 right, besides to achieve a total different scale for both the devices, appear to be driven by a different gradient of better performance in each device. This

Table 32 – Table presenting the mean and standard deviation value, with respect to the Device B, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the first step belonging to the standard operation mode, portrayed by the tables 15 and 16.

Trial number	ASSD mean x-axis	Std x-axis	ASSD mean y-axis	Std y-axis
1	0.0503	0.0022	0.1043	0.0102
2	0.0677	0.0035	0.2608	0.0035
3	0.0283	0.0021	0.4041	0.1757
4	0.3682	0.0517	1.9824	0.0902
6	0.0793	0.0050	0.1937	0.0072
5	0.1354	0.0042	0.6015	0.0030
7	0.0517	0.0196	1.4077	0.0546
8	1.1927	0.0882	6.1877	0.1071

Figure 52 – Figure presenting the performances of two inclinometer devices. The experiments were carried out through the exact same systematic performance from Section 3.4, so that both are analysed under the same conditions. Both the devices perform measurements according to the volume of parameters regarded by the Table 17, which concerns the second step of the standard operation mode's DOE. The device A refers to the device that was employed in this work. All the experiments of the DOE and machine learning approach were carried out by the device A. The device B was just employed in this work in order to compare the responses with the device A.



Source: Author.

last statement can be remarked when analysing the trials number 6 and 7 of the sensing y-axis, where the ASSD value increases from trial 6 to 7, for the device B, and decrease for the device A. The Table 32 exposes the values of the mean value of ASSD results regarding the three

conjunct of trial's performances for the Device B. The results of this table can be compared with the results, for the same conditions, of the Device A, in Table 25. The Device's B variability, assessed through the standard deviation (abbreviation Std in Table 32) of the respective trials from each conjunct of trials, when compared with the Device's A variability (also in Table 25), presents itself roughly similar for the sensing x-axis, but totally different for the sensing y-axis. An interesting observation concerns the quality of performance for the two devices in comparison. In Figure 51, one may remark that the Device B achieves smaller ASSD values than the Device A, roughly for all the trials numbers regarding the sensing x-axis. The opposite occurs for the sensing y-axis, where the Device A achieves smaller ASSD values than the Device B.

These results comparing the performance of the standard operation mode's first step for two devices, convey the knowledge that the Device B has a totally distinct behaviour in comparison with the device A, for most of the trials numbers of the volume of parameters from Table 15, designed in Table 16. These results also convey that, if the concept of replication were deployed using two sensors devices in the DOE architecture, the average response of the ASSD values given by both the devices would not be a reliable notion for neither the devices.

In order to remark whether the devices' different behaviours would remain for another volume of parameters, another step of experiments was carried out with the Device B, namely the second step of the standard operation mode's DOE. The Figure 52 and Table 33 present the responses of both the Devices A and B, for the set of parameters specified. One may remark that, the responses of the Device A are the very same found in Subsubsection B.3.1 for the second step.

Table 33 – Table presenting the mean and standard deviation value, with respect to the Device B, regarding the sensing axis x and y, of the three conjunct of trials' ASSD responses from the second step belonging to the standard operation mode, portrayed by the tables 17 and 16.

Trial number	ASSD mean x-axis	Std x-axis	ASSD mean y-axis	Std y-axis
1	0.0526	0.0016	0.0975	0.0101
2	0.0527	0.0032	0.0910	0.0029
3	0.0421	0.0007	0.0800	0.0069
4	0.0422	0.0003	0.0652	0.0034
5	0.0614	0.0019	0.1208	0.0118
6	0.0641	0.0021	0.1025	0.0046
7	0.0457	0.0020	0.1206	0.0034
8	0.0491	0.0018	0.0790	0.0043

In Figure 52 left, which stands for the responses from the sensing x-axis, one may note that, differently from the responses of the previous step, both the devices achieved akin ASSD values. Again in contrast, the sensing y-axis' responses, Figure 52 right, were found drastically different for each device. The device B achieved ASSD values much higher than the Device A.

One may note that, in the previous step, the Device B had achieved better performances than the Device A. This relationship appears to had been changed through the outer to inner volume of parameters. The Table 33 presents the variability, assessed through the standard deviation, of the respective trials numbers from the three conjunct of trials that were carried out in this second step of the Device B. One may note that, the variability of the trials numbers concerned the Device B, when compared to the respective from the Device A in Table 26, is quite similar, though slightly greater.

The Device's B second step of experiments responses complies with the assertions that the Devices A and B are driven to different performances by the same set of parameter. Such a fact justifies the reason to do not include the Device B into the DOE experiments and the machine learning approach. Some hypothesis can be suggested to explain the device's different performance behaviour. Since the inclinometer device is handmade, imperfections in the component's soldering are inevitable. Such imperfections vary from device to device, and could justify singularities for each inclinometer. There is also possible that imperfections in the electronic components could lead to different responses among the sensors. At the time of this work, no other devices were able to be compared.

C APPENDICE - STANDARD OPERATION MODE'S DATASET

Table 34 – Table presenting the train set employed to train the machine learning model of the standard operation mode.

Trial	Interval	Frequency	nValue	Timesleep	OM	ASSD X	ASSD Y
1	400	5	10	0.4	STANDARD	1.9787	2.1555
2	400	5	10	0.7	STANDARD	0.3910	0.2472
3	400	5	10	1.0	STANDARD	0.3353	0.0981
4	400	5	10	1.3	STANDARD	0.3343	0.0929
5	400	30	10	0.4	STANDARD	0.4100	0.4610
6	400	30	10	0.7	STANDARD	0.2477	0.0542
7	400	30	10	1.0	STANDARD	0.0718	0.0386
8	400	30	10	1.3	STANDARD	0.0622	0.0381
9	400	70	10	0.4	STANDARD	0.2702	0.3053
10	400	70	10	0.7	STANDARD	0.0766	0.0468
11	400	70	10	1.0	STANDARD	0.0663	0.0402
12	400	70	10	1.3	STANDARD	0.0604	0.0392
13	400	120	10	0.4	STANDARD	0.2381	0.2317
14	400	120	10	0.7	STANDARD	0.0773	0.0434
15	400	120	10	1.0	STANDARD	0.0741	0.0406
16	400	120	10	1.3	STANDARD	0.0732	0.0423
17	700	5	10	0.4	STANDARD	1.3063	1.5127
18	700	5	10	0.7	STANDARD	0.2242	0.1330
19	700	5	10	1.0	STANDARD	0.2373	0.0758
20	700	5	10	1.3	STANDARD	0.2354	0.0740
21	700	30	10	0.4	STANDARD	0.2353	0.2670
22	700	30	10	0.7	STANDARD	0.2004	0.1628
23	700	30	10	1.0	STANDARD	0.0501	0.0287
24	700	30	10	1.3	STANDARD	0.0358	0.0275
25	700	70	10	0.4	STANDARD	0.1573	0.1810
26	700	70	10	0.7	STANDARD	0.0474	0.0326
27	700	70	10	1.0	STANDARD	0.0422	0.0313
28	700	70	10	1.3	STANDARD	0.0439	0.0318
29	700	120	10	0.4	STANDARD	0.1456	0.1372
30	700	120	10	0.7	STANDARD	0.0506	0.0326
31	700	120	10	1.0	STANDARD	0.0486	0.0375

32	700	120	10	1.3	STANDARD	0.0511	0.0335
33	1000	5	10	0.4	STANDARD	0.8267	0.8394
34	1000	5	10	0.7	STANDARD	0.1269	0.0572
35	1000	5	10	1.0	STANDARD	0.1120	0.0571
36	1000	5	10	1.3	STANDARD	0.0430	0.0467
37	1000	30	10	0.4	STANDARD	0.1721	0.1711
38	1000	30	10	0.7	STANDARD	0.0389	0.0260
39	1000	30	10	1.0	STANDARD	0.0372	0.0258
40	1000	30	10	1.3	STANDARD	0.0296	0.0252
41	1000	70	10	0.4	STANDARD	0.1157	0.1204
42	1000	70	10	0.7	STANDARD	0.0424	0.0296
43	1000	70	10	1.0	STANDARD	0.0407	0.0289
44	1000	70	10	1.3	STANDARD	0.0398	0.0310
45	1000	120	10	0.4	STANDARD	0.1036	0.0950
46	1000	120	10	0.7	STANDARD	0.0483	0.0306
47	1000	120	10	1.0	STANDARD	0.0457	0.0303
48	1000	120	10	1.3	STANDARD	0.0442	0.0293
49	1300	5	10	0.4	STANDARD	0.6795	0.6845
50	1300	5	10	0.7	STANDARD	0.1105	0.0497
51	1300	5	10	1.0	STANDARD	0.1213	0.0564
52	1300	5	10	1.3	STANDARD	0.0859	0.0423
53	1300	30	10	0.4	STANDARD	0.1345	0.1250
54	1300	30	10	0.7	STANDARD	0.0336	0.0242
55	1300	30	10	1.0	STANDARD	0.0325	0.0241
56	1300	30	10	1.3	STANDARD	0.0315	0.0233
57	1300	70	10	0.4	STANDARD	0.0924	0.0881
58	1300	70	10	0.7	STANDARD	0.0376	0.0277
59	1300	70	10	1.0	STANDARD	0.0360	0.0228
60	1300	70	10	1.3	STANDARD	0.0348	0.0250
61	1300	120	10	0.4	STANDARD	0.0849	0.0669
62	1300	120	10	0.7	STANDARD	0.0412	0.0251
63	1300	120	10	1.0	STANDARD	0.0391	0.0251
64	1300	120	10	1.3	STANDARD	0.0405	0.0271

Table 35 – Table presenting the test set employed to assess the machine learning model for the standard operation mode.

Trial	Interval	Frequency	nValue	Timesleep	OM	ASSD X	ASSD Y
1	550	18	10	0.55	STANDARD	0.4270	0.3789
2	550	50	10	0.55	STANDARD	0.2217	0.1801
3	550	95	10	0.55	STANDARD	0.1862	0.1618
4	850	18	10	0.55	STANDARD	0.2625	0.2090
5	850	50	10	0.55	STANDARD	0.1506	0.1179
6	850	95	10	0.55	STANDARD	0.1245	0.0948
7	1150	18	10	0.55	STANDARD	0.2016	0.1627
8	1150	50	10	0.55	STANDARD	0.1092	0.0910
9	1150	95	10	0.55	STANDARD	0.0984	0.0592
10	550	18	10	0.85	STANDARD	0.0486	0.0492
11	550	50	10	0.85	STANDARD	0.0516	0.0434
12	550	95	10	0.85	STANDARD	0.0597	0.0434
13	850	18	10	0.85	STANDARD	0.0379	0.0354
14	850	50	10	0.85	STANDARD	0.0494	0.0359
15	850	95	10	0.85	STANDARD	0.0623	0.0367
16	1150	18	10	0.85	STANDARD	0.0378	0.0340
17	1150	50	10	0.85	STANDARD	0.0438	0.0341
18	1150	95	10	0.85	STANDARD	0.0529	0.0367
19	550	18	10	1.15	STANDARD	0.0505	0.0475
20	550	50	10	1.15	STANDARD	0.0535	0.0433
21	550	95	10	1.15	STANDARD	0.0666	0.0439
22	850	18	10	1.15	STANDARD	0.0423	0.0358
23	850	50	10	1.15	STANDARD	0.0501	0.0349
24	850	95	10	1.15	STANDARD	0.0593	0.0378
25	1150	18	10	1.15	STANDARD	0.0347	0.0342
26	1150	50	10	1.15	STANDARD	0.0409	0.0333
27	1150	95	10	1.15	STANDARD	0.0492	0.0344

D APPENDICE - REPEATED OPERATION MODE'S DATASET

Table 36 – Table presenting the train set employed to train the machine learning model of the repeated operation mode.

Trial	Interval	Frequency	nValue	OM	ASSD X	ASSD Y
1	300	7	10	REPEATED	0.0754	0.0784
2	300	20	10	REPEATED	0.0570	0.0530
3	300	40	10	REPEATED	0.0740	0.0525
4	300	60	10	REPEATED	0.0818	0.0541
5	300	80	10	REPEATED	0.0874	0.0588
6	300	100	10	REPEATED	0.0942	0.0543
7	300	120	10	REPEATED	0.1007	0.0552
8	500	7	10	REPEATED	0.0614	0.0652
9	500	20	10	REPEATED	0.0491	0.0407
10	500	40	10	REPEATED	0.0635	0.0427
11	500	60	10	REPEATED	0.0655	0.0444
12	500	80	10	REPEATED	0.0726	0.0433
13	500	100	10	REPEATED	0.0757	0.0464
14	500	120	10	REPEATED	0.0831	0.0470
15	700	7	10	REPEATED	0.0550	0.0557
16	700	20	10	REPEATED	0.0453	0.0370
17	700	40	10	REPEATED	0.0528	0.0401
18	700	60	10	REPEATED	0.0581	0.0394
19	700	80	10	REPEATED	0.0579	0.0374
20	700	100	10	REPEATED	0.0649	0.0418
21	700	120	10	REPEATED	0.0665	0.0407
22	900	7	10	REPEATED	0.0473	0.0479
23	900	20	10	REPEATED	0.0378	0.0361
24	900	40	10	REPEATED	0.0467	0.0352
25	900	60	10	REPEATED	0.0527	0.0423
26	900	80	10	REPEATED	0.0554	0.0361
27	900	100	10	REPEATED	0.0606	0.0385
28	1100	40	10	REPEATED	0.0441	0.0334
29	1100	60	10	REPEATED	0.0465	0.0374
30	1100	80	10	REPEATED	0.0521	0.0330
31	1100	100	10	REPEATED	0.0518	0.0346

32	1100	120	10	REPEATED	0.0568	0.0378
33	1300	7	10	REPEATED	0.0448	0.0345
34	1300	20	10	REPEATED	0.0336	0.0336
35	1300	40	10	REPEATED	0.0428	0.0332
36	1300	60	10	REPEATED	0.0461	0.0301
37	1300	80	10	REPEATED	0.0507	0.0332
38	1300	100	10	REPEATED	0.0497	0.0363
39	1300	120	10	REPEATED	0.0535	0.0343
40	1500	7	10	REPEATED	0.0375	0.0341
41	1500	20	10	REPEATED	0.0324	0.0311
42	1500	40	10	REPEATED	0.0373	0.0333
43	1500	60	10	REPEATED	0.0453	0.0347
44	1500	80	10	REPEATED	0.0426	0.0320
45	1500	100	10	REPEATED	0.0493	0.0279
46	1500	120	10	REPEATED	0.0489	0.0345

Table 37 – Table presenting the test set employed to assess the machine learning model of the repeated operation mode.

Trial	Interval	Frequency	nValue	OM	ASSD X	ASSD Y
1	400	13	10	REPEATED	0.0482	0.0562
2	400	30	10	REPEATED	0.0537	0.0424
3	400	50	10	REPEATED	0.0659	0.0483
4	400	70	10	REPEATED	0.0695	0.0448
5	400	90	10	REPEATED	0.0780	0.0475
6	400	110	10	REPEATED	0.0858	0.0462
7	600	13	10	REPEATED	0.0439	0.0396
8	600	30	10	REPEATED	0.0478	0.0341
9	600	50	10	REPEATED	0.0537	0.0383
10	600	70	10	REPEATED	0.0561	0.0379
11	600	90	10	REPEATED	0.0600	0.0401
12	600	110	10	REPEATED	0.0578	0.0395
13	800	13	10	REPEATED	0.0401	0.0372
14	800	30	10	REPEATED	0.0404	0.0398
15	800	50	10	REPEATED	0.0450	0.0384
16	800	70	10	REPEATED	0.0495	0.0377
17	800	90	10	REPEATED	0.0499	0.0403
18	1000	50	10	REPEATED	0.0421	0.0423
19	1000	70	10	REPEATED	0.0475	0.0382
20	1000	90	10	REPEATED	0.0479	0.0432
21	1200	50	10	REPEATED	0.0412	0.0395
22	1200	70	10	REPEATED	0.0412	0.0379
23	1200	90	10	REPEATED	0.0479	0.0367
24	1200	110	10	REPEATED	0.0474	0.0407
25	1400	13	10	REPEATED	0.0340	0.0383
26	1400	30	10	REPEATED	0.0365	0.0395
27	1400	50	10	REPEATED	0.0390	0.0370
28	1400	70	10	REPEATED	0.0416	0.0405
29	1400	90	10	REPEATED	0.0426	0.0418
30	1400	110	10	REPEATED	0.0474	0.0401