



**Politecnico
di Torino**

Politecnico di Torino

Corso di Laurea in Mechanical Engineering

A.a. 2022/2023

Sessione di Laurea Dicembre 2023

Development of a simplified SOFC model using Machine Learning

Relatori:

Prof. Vittorio Verda

Candidati:

Andrea Natali (s282171)

Corelatori:

Prof.ssa Martina Capone

Dott. Ing. Andreas Schulze

“Homo faber fortunae suae”
Appio Claudio Cieco (Sententiae)

Table of Contents

INTRODUCTION	1
1. FUEL CELLS OVERVIEW	4
1.1 FUEL CELL TYPES.....	5
1.2 ENERGY CONVERSION EFFICIENCY	6
1.3 APPLICATIONS AND LIMITS	7
1.4 Solid Oxide Fuel Cell - SOFC	8
1.4.1 SOFC WORKING PRINCIPLE.....	8
1.4.2 MATERIALS FOR SOFCs	9
2. ENGINEERING MODELING	12
2.1 TYPES OF MODELING.....	12
2.1.1 WHITE-BOX MODELING	13
2.1.2 BLACK BOX MODELING	14
2.1.3 GRAY BOX MODELING	19
3. SOFC PHYSICAL MODEL.....	21
3.1 CELL VOLTAGE.....	21
3.2 ACTIVATION LOSSES	23
3.3 OHMIC LOSSES.....	24
3.4 CONCENTRATION LOSSES	25
3.5 DISCUSSION ON THE PHYSICAL MODEL.....	27
4. SOFC SIMULATION MODEL	29
4.1 GEOMETRY	29
4.2 PHYSICS AND BOUNDARY CONDITIONS	30
4.2.1 CATHODE DOMAIN	30
4.2.2 ELECTROLYTE DOMAIN.....	32
4.2.3 ANODE DOMAIN	33

4.2.4 SUBSTRATE DOMAIN	33
4.3 WORKFLOW AND EXPECTED RESULTS	34
5. SENSITIVITY ANALYSIS	36
5.1 PRELIMINARY PARAMETER SELECTION	36
5.2 PROCEDURE.....	38
5.3 RESULTS	38
6. SOFC GREY-BOX MODEL.....	42
6.1 SIMULATION DATASET	42
6.2 EVALUATIONS	43
6.2.1 HUMIDIFIED HYDROGEN MODEL	44
6.2.2 HYDROCARBON BASED MODEL	47
6.3 COMMENTS.....	50
7. SOFC BLACK-BOX MODELS.....	51
7.1 LINEAR MODELS	51
7.1.1 LINEAR REGRESSION	52
7.1.2 BAYESIAN RIDGE REGRESSION	55
7.1.3 POLYNOMIAL REGRESSION	56
7.2 CLASSIFICATION ALGORITHMS.....	62
7.2.1 DECISION TREES.....	62
7.2.2 GRADIENT BOOSTING.....	64
7.2.3 RANDOM FOREST.....	67
7.2.4 HISTOGRAM-BASED GRADIENT BOOSTING.....	69
7.3 ARTIFICIAL NEURAL NETWORKS.....	71
7.4 RESULTS	76
7.5 MODEL APPLICABILITY	77
CONCLUSIONS	80

INTRODUCTION

While the Automotive industry is pushing towards the complete electrification of cars and the classic means of transport to reduce the emissions of harmful gases in the atmosphere, the actual electricity demand, and the larger upcoming one, is still satisfied in its majority by fossil fuels. According to the available data provided by *Ember's Yearly Electricity Data*, in 2022 behind the 28527.76 TWh of globally produced electricity, almost 61% was covered by fossil fuels such as Coal, Gas and Oil [1]. If we consider the forecasting provided by *McKinsey & Co.* on *Statista.com*, global electricity production is expected to increase by 296.5% by 2050 (83000 TWh), [2] the same year fixed as a target for Carbon Neutrality by the European Union.

Given these assumptions, scientists' efforts moved towards more efficient and more environmentally friendly sources of energy. Fuel cells are a highly efficient technology for various applications because they generate electricity using a range of fuels, including both traditional and eco-friendly options, through electrochemical processes. This method allows them to avoid the limitations of the Carnot cycle and produce environmentally friendly products. Consequently, fuel cells offer higher efficiency compared to conventional power generation methods. The fields of possible applications are countless, but considering just the power generation purposes, among the different types of fuel cells, high-temperature fuel cells (SOFCs, MCFCs) are the most efficient devices and are recently being adopted in the clean distributed generation systems. [3]

The SOFC R&D department in *Robert Bosch GmbH* gave me the opportunity to work on this product fascinating and complex at the same time. In this stimulating environment, I had the chance to deepen engineering modeling and work on a possible

game changer for green energy production in the near future. This collaboration had the scope to extend a physical model, valid in a certain dimension range of the cell, towards a larger scale that resulted impossible to simulate via conventional software and methods. Due to the impossibility of working with white/grey box models and the non-linearity of the physics, it was chosen to try via a black-box modeling approach to predict the behavior and create a complete mapping of the interesting properties across the single repetitive unit.

In this study, the development started from the physical model, used to describe the single-cell functionality, already implemented and validated through a simulation model in *COMSOL Multiphysics*. The latter was investigated to retrieve the most important features among all the actors through a Sensitivity Analysis carried out using statistical tools such as the Design of Experiments techniques. Finally, the model was extended to a larger scale via the predictions performed by different Machine Learning algorithms that will be compared in performance and accuracy. The most accurate algorithm resulted in the *Histogram-based Gradient Boosting Regressor*; the latter was trained and implemented in a web application with the scope to create a user-friendly prediction tool.

1. FUEL CELLS OVERVIEW

Fuel cells are devices that convert the chemical power of a fuel utilizing electron and ion exchange through an electrolyte, the latter is generated by reduction-oxidation (RedOx) reactions that take place respectively at the cathode and at the anode of the cell. The principle of operation behind fuel cells can be dated to the early 1800s when Mr. *Henry David* announced the principle of reverse electrolysis highlighting the possibility to generate electricity by the collection of electrons from the reaction between oxygen and hydrogen. *Sir William Grove* is widely considered the inventor of modern fuel cells since he built, tested and demonstrated the first cell stack called “voltaic gas battery” back in 1839. His work was then continued and deepened by personalities such as *Mond* and *Langer* who first used a solid porous non-conductive diaphragm impregnated in acid before being plated and placed in contact with the two working gases. This solution was the precursor of modern fuel cells that have a structure very similar to that presented by the scientists in 1889. Ten years later, in 1899, *William Nernst* first used zirconium oxide for the filament of his light bulb after carrying out studies on the solid-ionic conductivity of stabilized zirconia (zirconium oxide doped with few moles per cent of calcia, magnesia or yttria). As a result of his studies, he took advantage of the phenomena of their increasing conductivity with rising temperature, creating a very efficient light bulb with respect to the pre-existent carbon fiber bulb. By the end of the 1930s, *Baur* and *Preis* used doped zirconia to build fuel cells and developed the two designs of cells that are still adopted today: Tubular-shaped and Planar fuel cells. [3]

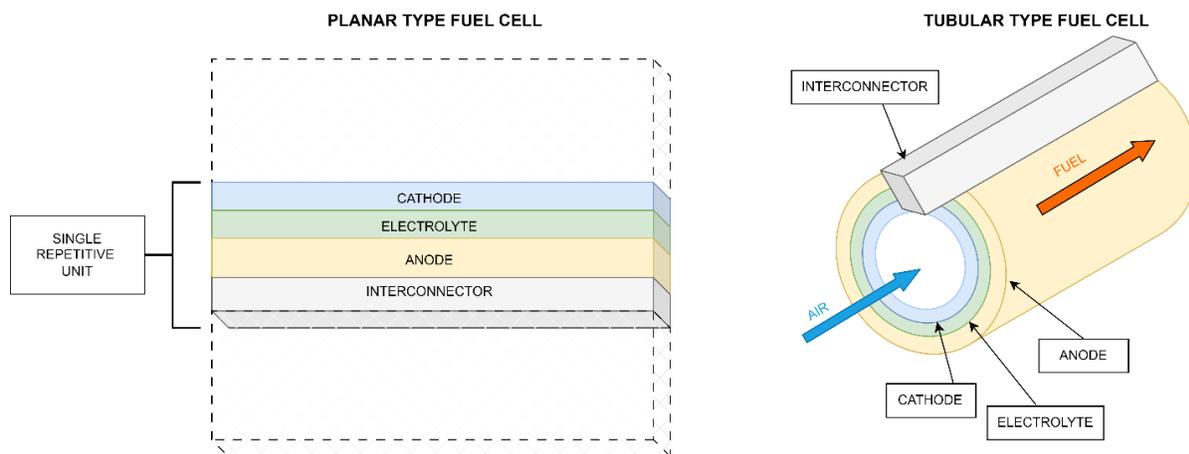


Figure 1 - Planar type (A) and Tubular type (B) fuel cells

Although changing the shape, all fuel cells are composed of the following elements:

- **Electrodes:** Anode and Cathode, both usually made of metals. The anode is placed in contact with the fuel and the oxidation reaction takes place. The cathode is in contact with air to allow the reduction reaction to take place.
- **Electrolyte:** it is a key component whose main purpose is to physically separate the anode (fuel side) from the cathode (air side) guaranteeing electrical insulation and the transport of ions (or protons) according to the reaction kinetics. The material of the electrolyte can be different and define the type of fuel cell.
- **External circuit:** it is an electrical circuit implemented in the design to collect and transfer the electrons outside the cell.
- **Catalysts:** both for the anode and the cathode. They are important components that enhance the efficiency and kinetics of the reactions that take place on the electrodes' surfaces.

1.1 FUEL CELL TYPES

Currently, the various types of fuel cells can be characterized by the material of the electrolyte adopted:

- Alkaline Fuel Cell – *AFC*
- Direct Methanol Fuel Cell – *DMFC*
- Molten Carbonate Fuel Cell – *MCFC*
- Phosphoric Acid Fuel Cell – *PAFC*
- Proton Exchange Membrane Fuel Cell – *PEMFC*
- Solid Oxide Fuel Cell – *SOFC*

Electrolyte type and material strongly influence some characteristics like the side of the fuel cell where the main reactions occur and where the reaction products are exhausted (anode or cathode). But the most important property is the conductivity which is strictly related to the operational temperature. Therefore, even if different types of cells can reach the same conductivity, the operational temperature may vary quite largely [3]. In *Table 1* can be grouped the main characteristics of the different types of fuel cells:

	PEMFC	PAFC	MCFC	SOFC
<i>Electrolyte</i>	Polymeric membrane	Phosphoric Acid	Liquid Carbonate	Solid Oxide
<i>Temperature °C</i>	80	200	650	800-1000
<i>Efficiency %</i>	40-50	40-50	>60	>60
<i>Ions carrier</i>	Hydrogen protons	Hydrogen protons	Carbonate ions	Oxygen ions
<i>Catalysts</i>	Platinum	Platinum	Nickel	Perovskite
<i>Fuel</i>	H ₂	H ₂	H ₂ , CO, CH ₄	H ₂ , CO, CH ₄
<i>Poisons</i>	CO	CO, Sulfur, H ₂ S	Sulfur, H ₂ S	Sulfur, H ₂ S
<i>State of development</i>	Pre-commercial	Commercial	Prototype	Pre-commercial

Table 1 - Fuel cell types comparison [3]

1.2 ENERGY CONVERSION EFFICIENCY

The Energy Conversion Efficiency of a conversion device or process is uniquely defined as [4]:

$$\eta_{EC} = \frac{\text{Output Energy}}{\text{Input Energy}} = \frac{\text{Electrical Energy}}{\text{Chemical Energy of Fuel}}$$

Generally speaking, fuel cells are very efficient devices since they can directly transform the chemical power of the fuel into electrical energy via electrochemical reactions. The traditional power generation plants (excluding hydraulic and nuclear plants) usually require the combustion of fossil fuels to power mechanical systems that transform mechanical energy into electrical one. The combustion process itself is a very practical and effective way to power mechanical devices such as turbomachines or ICEs. Nevertheless, the overall efficiency does not exceed 40%, due to the large amount of energy converted and wasted into heat. Fuel cells, which are not related to the Carnot Cycle's limitations and are not paired to other mechanical components, are more likely to have efficiencies that range from 35 to 60% considering just the energy conversion. The high-temperature cells, such as MCFCs and SOFCs, can also be exploited to build hybrid systems (when connected to traditional plants) or in combined

Heat&Power configurations that push the overall efficiency up to 85% [3]. However, the efficiency of these devices is strictly related to several factors that involve the need to adopt complex designs and a fine-tuning of the operating conditions. Some of the agents affecting the performances are listed below [5]:

- Catalysts Efficiency
- Fuel Utilization (FU)
- Operating Temperature
- Ionic Conductivities of the materials
- Gas Diffusion and Mass Transport properties of the materials
- Leakages of gas species and current
- Purity of the gas species (i.e. Fuel, Air)
- Stack Design and Configuration

1.3 APPLICATIONS AND LIMITS

Fuel cells are devices, that, compared to the classical solutions for power generation, directly transform the chemical power of the fuel into electricity without being paired to mechanical systems thus guaranteeing higher energy conversion efficiency. When the cells are fueled with hydrogen the main byproduct of the reaction is water vapor thus making these devices a clean energy technology, while if we consider all the possible fuels that can be employed (like methanol, biofuels, natural gas or hydrogen-rich gases) the products have only very small percentages of harmful or greenhouse gases especially when compared with the classic alternatives. Considering the advantages listed above, one of the most popular applications in the future will be the Distributed Generation (DG). The research on solutions for distributed generation arises from the need to solve problems related to the use of centralized energy systems such as the need to predict the annual and daily demand, the use of complex control systems in the power grid for the management of consumers and the impossibility of free and flexible access to resources by users. High-temperature fuel cells (SOFCs, MCFCs) in particular, allow for building systems of energy distribution where the production is made locally and the role of the transmission network is eliminated or greatly reduced, thus leading to wide flexible ranges of power with higher efficiency, higher network efficiencies and maintenance-free operation. These devices also guarantee acceptable operation costs while adding the possibility to use the output

thermal energy by conveying it to heating systems, increasing the overall efficiency of the system. Countless other applications see fuel cells used in sectors such as the automotive, aerospace and technology industries [3] [6].

However, this technology still has several limitations that will be listed below:

- **High Installation costs:** although operating and fuel costs can be compared to the classical solutions for power generation, the installation cost is still the main barrier to their commercialization and it is due to the expensive materials used for catalysts (e.g. Platinum, Nickel).
- **Hydrogen-related costs:** given the lack of adequate infrastructure and the problems related to production, transport and storage, the difference in costs is considerable when compared to conventional fuels.
- **Durability and reliability:** the studies carried out on the plants installed in the past decades, have shown a certain sensitivity to degradation over time particularly when used in difficult operating conditions such as high temperature and chemical poisoning [4].

Although the advantages are greater than the disadvantages, research in this field must be pushed forward to enable cost reduction and greater economic benefits for customers. Furthermore, must be enhanced the development of adequate infrastructure to make access to hydrogen easy and affordable for everyone.

1.4 Solid Oxide Fuel Cell - SOFC

This chapter aims to analyze more in depth the Solid Oxide Fuel Cells as this thesis work was conducted on a single cell of the SOFC engineered by the team of *Robert Bosch GmbH* in collaboration with *Ceres Power*. Understanding the working principles and the materials used is fundamental for the context of this master's thesis which focuses on the modeling and the prediction of the cell's behavior in terms of performance.

1.4.1 SOFC WORKING PRINCIPLE

The singular cell is composed of three different layers: anode, cathode and electrolyte. The cathode is one of the two porous electrodes present in the cell. It works in an oxidizing environment with temperatures from 600-1000°C and is responsible for the

production of oxygen ions through the electrochemical reaction of oxygen, starting from the gaseous phase to oxide ions consuming two electrons. The oxygen ions are then incorporated into the electrolyte's oxygen vacancies to reach the anode. The latter is the second porous electrode of the cell and catalyzes the reaction between the fuel (pure hydrogen or hydrogen derived from steam reforming processes) and oxygen ions producing water and two electrons as byproducts. The electrons produced are then collected into the interconnector which is used to connect multiple single repetitive units into the so-called "cell stack".

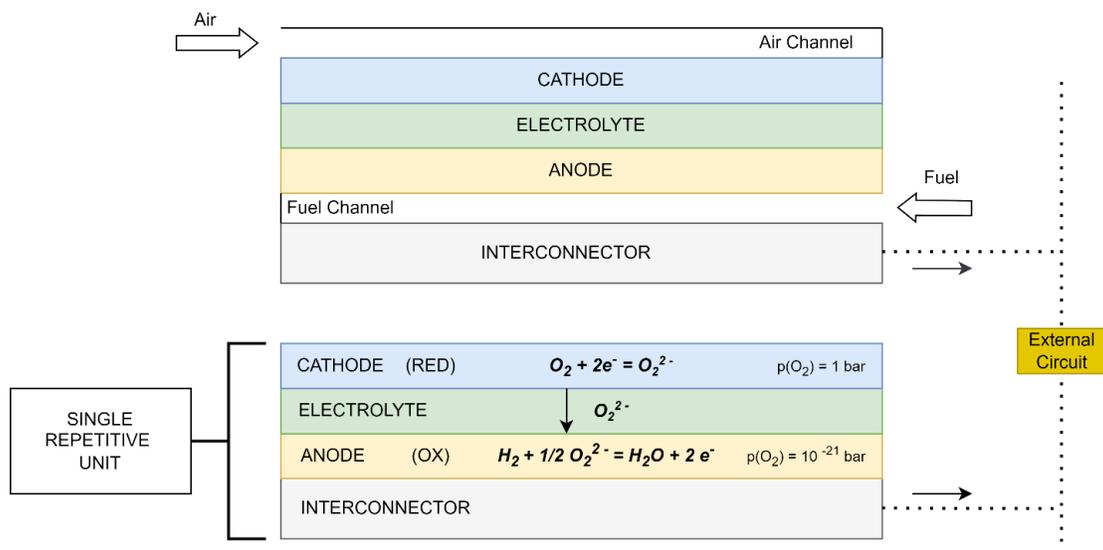


Figure 2 - Scheme of a singular fuel cell and main reactions

1.4.2 MATERIALS FOR SOFCs

The materials that compose the different elements of the SOFCs play a key role in the cells' performance and are an active component in the reactions' kinetics. The goal of this paragraph is to underline the main characteristics required for each material, based on the role it plays in the cell.

1.4.2.1 CATHODE

SOFC cathode materials must be catalytically active for oxygen reduction furthermore, they must be able to bear a constant oxidizing atmosphere with temperature that ranges

from 600 to 1000°C. Other important requirements for the correct air electrode functioning are listed below:

1. **High electronic conductivity:** the cathode material must be able to provide pathways for the electrons between the reaction site and the external circuit.
2. **High ionic conductivity:** it is required to facilitate the transport of oxygen anions towards the electrolyte contact interface.
3. **Chemical and dimensional stability:** these are important characteristics to be met to guarantee the correct production and functioning during operation.
4. **Sufficient porosity:** fundamental to obtain the transport of the gas phase to the electrolyte interface.
5. **Low activation overpotential:** to reduce the activation losses.

The suitable materials that meet the requirements above are mostly oxide-based, therefore Doped Perovskite oxides (i.e. LSM, LSCFO, LSCM) are adopted for this purpose since they guarantee good mixed conductivity as well as sufficient thermal stability and compatibility with the other interfacing materials [7].

1.4.2.2 ELECTROLYTE

The electrolyte has the main role of conducting the anions from the cathode to the anode, without proper conduction properties the voltage generated by the cell would be too low to be exploited. The SOFCs use an oxide-based electrolyte in the solid phase which must guarantee a porous-free layer between different gaseous phases. Other important characteristics that must be met are [7]:

1. **High ionic conductivity** at the desired operating temperature.
2. **Negligible electronic conductivity:** must isolate the anode from the cathode avoiding the short-circuiting of the system.
3. **Chemical and Mechanical compatibility.**
4. **Thermal stability and expansion compatibility:** must be stable at the desired temperature and have a thermal expansion compatible with the other actors.
5. **Easy fabrication into a dense and very thin membrane:** the lowest thickness and uniformity are fundamental to reduce ohmic losses.

For these purposes, the material that complies the most is the Ytria-Stabilized Zirconia (YSZ) which is the state of art of electrolyte currently used in SOFCs. However new

materials are being studied due to the key role that this layer has in the cell functionality such as perovskite oxides [7].

1.4.2.3 ANODE

The materials employed to produce the anode electrode must satisfy some requirements such as to be stable in a reducing environment and to be catalytically active for the fuel oxidation process [7]. Other requirements that must be met are:

1. **High electronic and ionic conductivity:** they must be able to receive the anions from the electrolyte and guarantee the flow of electrons and byproducts.
2. **Chemical and physical compatibility** with the other components.
3. **Sufficient porosity:** to reduce the losses associated with the diffusion of the species from the interface.
4. **Thermal and dimensional stability:** to be compatible with the expansion of the other materials.
5. **Easily produced in thin layers:** to reduce ohmic losses.

The materials suitable for these requirements are just a few types of composite materials made by metallic and ceramic phases called “cermet”. The metallic phase is usually fulfilled by Nickel or Nickel Oxide particles since Nickel is a cheap and efficient catalyzer for the division of the hydrogen atoms and guarantees a low carbon deposition, that causes clogging of the pores when hydrocarbons are used. The ceramic phase is made with Ytria-Stabilized Zirconia since it guarantees good ionic conductivity in the operating temperature range [7].

2. ENGINEERING MODELING

A *System* can be defined as: “[...] entirety of elements that relate with each other and interact in such a way that they can be regarded as a unit with a specific task, sense or purpose, and in this way differentiate themselves from the environment that surrounds them.” (BES-PE Glossary issue n.3-2009/11). According to this definition, a system can be imagined as a box filled with different actors that are interacting with each other, and the scientists or engineers are the observers that a looking from the outside. Most of the times the connections and the interactions between these objects are not immediately clear or simple to understand, therefore the need for another representation is somehow unavoidable. A *Model* can be defined as: “[...] representation of reality that is reduced to relevant characteristics.” (BES-PE Glossary issue n.3-2009/11). This definition anticipates one of the key points of the *System Analysis*, often referred to as *Modeling*: simplify the reality. The other reasons why scientists need to model a real system are: to acquire some knowledge and explain the behavior of the system through simulations (when studies of the real system are impossible or uneconomical) and to predict and optimize the system behavior before the system physically exists. The general procedure for modeling requires the following steps [8]:

1. **Creation of the model:** it is an iterative and heuristic process to adapt the model based on the knowledge of reality. The creation stops when the model itself can accomplish the task that was intended for.
2. **Validation:** it is based on the comparison between the observed properties and the predicted properties resulting from the model. If the predictions are correct the model can be considered “valid”, otherwise it will require more experiments to be validated.
3. **Simulation:** it is the implementation of the model for predicting the behavior of the system to acquire more knowledge.
4. **Further developing:** the model can be extended to different scales of reality, but it will require a new validating procedure.

2.1 TYPES OF MODELING

From the creation of the Scientific Method born by the genius of Galileo Galilei to the more contemporary Data Science, enormous steps have been carried out to allow more

efficient modeling, less time-consuming and adaptable to the more complex meanders of reality. Since the modeling procedure foresees the synergetic combination of experimentation and analysis, it therefore involves an expenditure in terms of money and time. Most of the time scientists and engineers face the need to find a suitable tradeoff between monetary resources and time expenditure in order to acquire a sufficient amount of knowledge to push forward their studies. Other times, they are forced to face real and physical limits that do not allow the analysis of a phenomenon through its pure observation with real experiments. Therefore, during the years and thanks to new technologies development, three main categories of modeling were born and are presented below:

- **White-Box Modeling**
- **Black-Box Modeling**
- **Gray-Box Modeling**

The intrinsic difference between categories can be summarized in *Figure 3*:

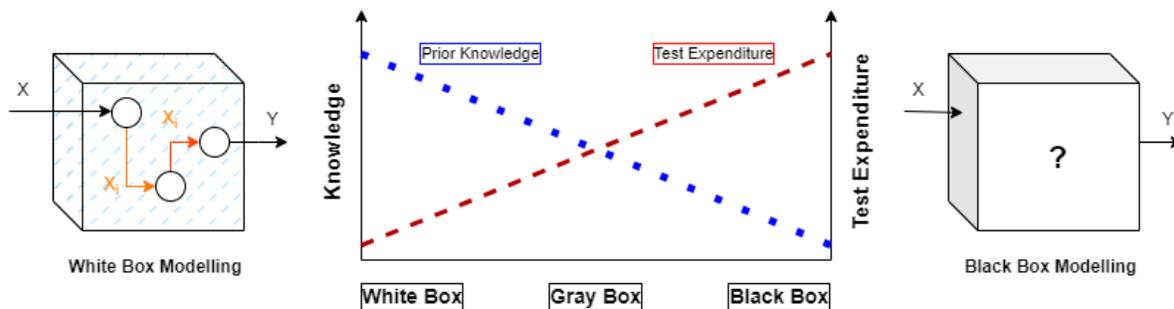


Figure 3 - Types of Modeling [8]

In the next subchapters, all the categories will be described, and some interesting tools will be explained more in detail for a better understanding of the following steps.

2.1.1 WHITE-BOX MODELING

White-Box Modeling, also called Physical Modeling, denotes a deductive process (from the general to the specific) on the basis of general natural laws and principles (e.g. laws of conservation, laws from chemistry, physics or thermodynamics), and structural knowledge of the system. Here the function of the system elements which correlates the output to the inputs, must be fully known in order to deduct the whole system behavior at the chosen level of observation. Model parameters must be defined on the basis of physical constants and very frequently the system can be represented as

a set of complex partial differential equations (*PDEs*). One of the strengths of this method is that the model can be applied over a wide spectrum of parameters since the applicability of the approach is only limited by its laws. Physical Modeling may be employed for technical questions if the state-of-the-art is adequate for the knowledge that needs to be acquired and for the task to be accomplished. Very often this approach does not apply to quantitative statements (i.e. knowledge retrieved from large amounts of data) since it will be time-consuming and not cost-effective. This approach will also lack precision if Variance affects the analyzed data, in that case, other types of approaches better suit the analyses. However qualitative analysis as well as the parameters interactions can be achieved through the study. [8]

2.1.2 BLACK BOX MODELING

Black box Modeling, also known as Empirical Modeling or System Identification, is an inductive approach (from the specific to the general) whereby a model is created only by recording the system inputs and outputs by way of an experiment in a finite number of discrete cases, e.g. through measurements and observation. This model then describes in the best possible way the relationships between inputs and outputs using a Transfer Function. The transfer function is calculated via a mathematical approach which usually involves a simple algebraic equation, e.g. a polynomial of the first or second order. Subsequent system identification is then achieved by determining the free model parameters (coefficients of the equation) so that the transfer function describes the system behavior within the framework of the model accuracy. The main advantage is that the procedure is usually faster because it does not require prior knowledge of the system's structure or the governing physical laws of the phenomena involved. Therefore, empirical modeling can provide the basis of physical modeling when the system is completely unknown and usually is used as the basis of the classical statistical experimental design (e.g. Factorial experimental designs). However, the model itself is valid inside narrow confines since the limited order of the model equation cannot be extended to the global system behavior. In that case, through Taylor Expansion, is possible to generalize the model up to a higher-order error [8].

2.1.2.1 DATA-DRIVEN MODELS

In the last thirty years, the technological evolution that involved the computer industry led to the complete digitalization of almost every aspect of our lives. To efficiently run these eco-systems of data-acquiring sensors the hardware of the computers (e.g. CPUs, RAMs, GPUs) was developed to such an extent that operations with a large amount of data and complexity can be performed in the order of microseconds. This evolution also touched the scientific world leading to the possibility of implementing mathematical algorithms to large amounts of data with a better fitting of the relationships between outputs and inputs, thus giving birth to Machine Learning (ML), Deep Learning (DL) and Artificial Intelligence (AI). These modeling approaches are defined as Data-driven models since they can retrieve patterns, relationships and knowledge from data they are trained on and perform predictions on different and unknown datasets (Supervised Learning). This procedure does not require any previous explicit programming with rules or assumptions [9].

In this dissertation different Machine Learning algorithms were implemented therefore a brief explanation is necessary for further understanding. Below is the classification of the ML algorithms used:

- **Linear Models:** These algorithms exploit Regression in order to correlate independent variables (also called “features”) to a dependent continuous output. The prediction power of this method is based on the degree of the polynomial used for the interpolation of the data. Polynomials with degree “d” equal to one are defined as “Linear Regression Models” while polynomials with a higher degree are defined generally as “Polynomial Regression Models”. An increase in the degree of the polynomial function at the base of the regression will produce a better fitting of the output data with the respective inputs since they will include also the interaction terms that arise from the polynomial construction. *Figure 4* below shows the difference in the data interpolation between a Linear Regression and a Polynomial Regression. Another Linear Model is the Bayesian Regression which adds a probabilistic approach (i.e. considers the uncertainties related to the features) to the regression by adding a prior distribution over the parameters (typically a Gaussian Distribution) and retrieve the probabilistic distribution of the results [10] [11] [12].

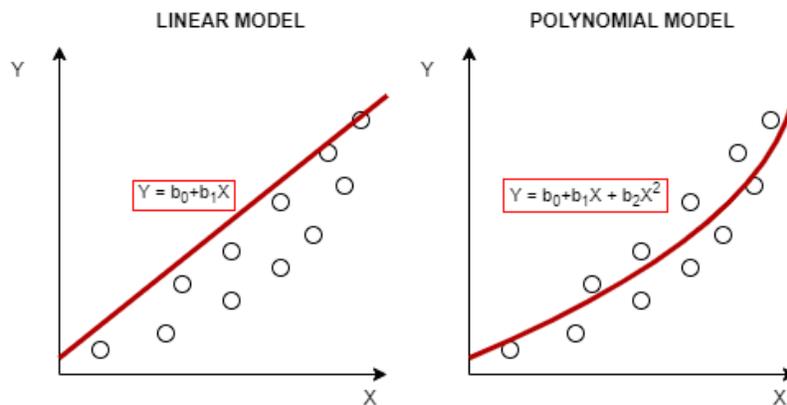


Figure 4 - Linear Regression ($d=1$) and Polynomial Regression ($d=2$) comparison

The Linear Models implemented in this thesis work are retrieved from the Python's library "Scikit-Learn" and can be recalled using their function's name as: *LinearRegression(degree = 1)* for linear regression, *LinearRegression(degree = 2, 3)* to call a polynomial regression, *Bayesian RegressionRidge* to call a Bayesian regression with a Ridge regression feature that allows to determine the best degree of the regressor and to deal with correlated inputs [11] [12] [13].

- Classification Models:** are Machine Learning models which use the principle of classification, i.e. a supervised algorithm trained to predict a correct "label" of a given input data. A "label" is a category that is assigned to the input data and will be used by the algorithm during the learning and the prediction procedure. The output of these models can be either discrete (e.g. binary, true/false, M/F) or continuous as for the linear models. If the output is continuous these models are considered meta-models since are still based on the regression of the data, but the decision is taken via categorized choices. The latter that can aim to a specific target such as the minimization of the residuals or the minimization of a loss function specifically designed for that algorithm. The simplest model is the *Decision Tree*, which is also called "weak learner" or "base learner" since it is the basic decision method employed in all the algorithms of this category. The intrinsic difference among all the algorithms is how the weak learner is arranged or shaped and how the decision is taken starting from the result of the decision tree implemented. As an example, *Random Forest* model implements multiple decision trees in parallel and takes

the decision basing on the average of the results of all the trees in the “forest”. The model can be tuned by defining some characteristics, known as “hyperparameters”, such as the number of trees in the forest, number of “leaves” on each tree, loss function to be averaged for the prediction. Other very common models are based on the *Gradient Boosting* method which involves the sequential (in series) training of base learners to fit the gradient of the loss function. Basically, the model assigns some “weights” to the input features and calculates the loss function of the error (e.g. residuals, mean squared error), then it finds the gradient of the function to know the direction of the maximum error. This optimization technique will induce the algorithm to set the parameters in such a way as to travel toward the opposite direction defined by the gradient until convergence is reached. In this way, the minimum of the loss function is searched thus resulting in more accurate predictions [14]. How the trees for different algorithms are managed is shown in the following *Figure 5*.

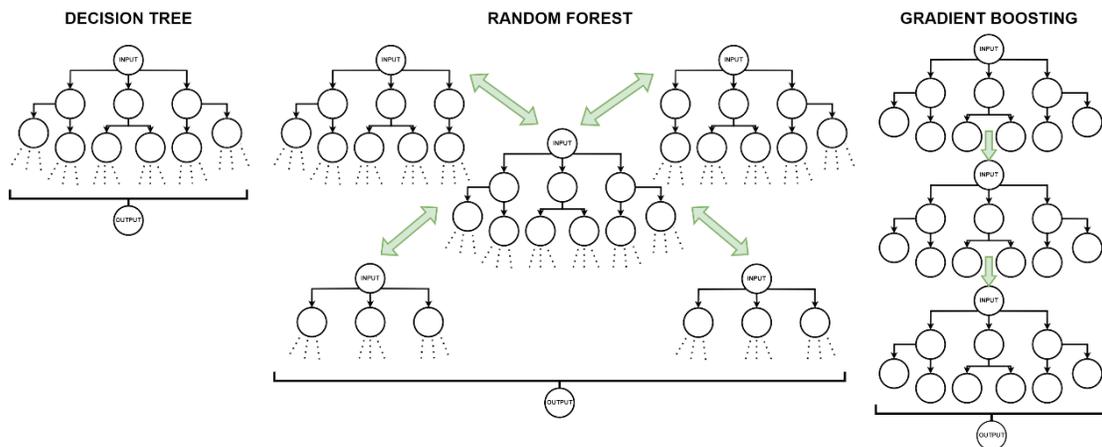


Figure 5 - Classification Models comparison

The classification algorithms treated in this dissertation are recalled from the Scikit-Learn library in Python and are: *DecisionTreeRegressor*, *RandomForestRegressor*, *GradientBoostingRegressor* and *HistGradientBoostingRegressor* [11] [12].

- Artificial Neural Networks:** are computational algorithms that are composed with a structure similar to the human brain network of neurons. Indeed, this model is characterized by the presence of nodes called “neurons” and is structured in layers, the first one is called the “input layer” and the last one is called the “output layer”. In between there are multiple layers called “hidden layers”, the number of hidden layers defines the type of the algorithm: if the number of hidden layers is between one and three the algorithm belongs to Machine Learning while, if this number is larger than three, the algorithm is considered as a Deep Learning model. The connection between the neurons is called “weight” and multiple neurons can be linked together. An important characteristic is the presence of an Activation Function that introduce the non-linearity therefore allowing the network to learn complex patterns, the latter is influenced by the input weights and can autonomously adapt to the nature of the problem. The parameters that can be controlled are the number of neurons in each layer, the number of hidden layers and the type of the activation function. Artificial Neural Network can be use either for classification, if the output is discrete, or for Regression if the output is continuous [3] [14]. Very unusual and powerful feature of this model is that the outputs can be multiple. *MLPRegressor* from *Scikit-Learn* was used for the analysis in the next chapters [11] [12]. The general structure of a Neural Network can be represented as in *Figure 6*:

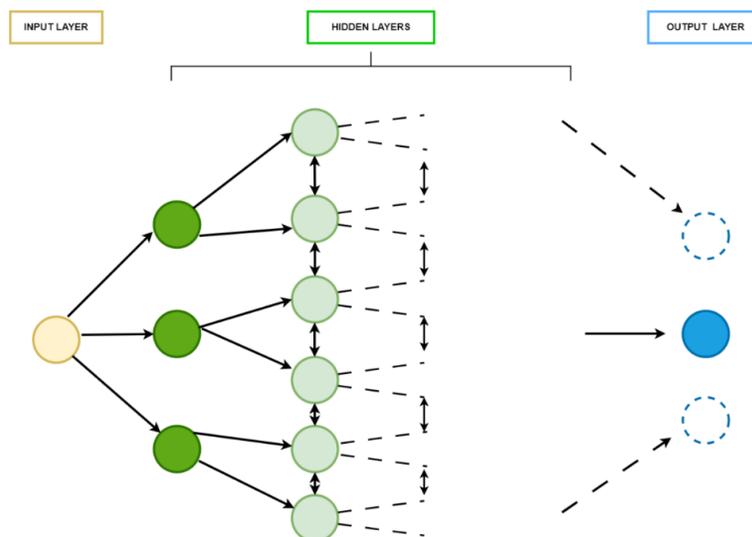


Figure 6 - Structure of an Artificial Neural Network

2.1.3 GRAY BOX MODELING

In addition to the two modeling approaches described above, various mixed forms exist, which come under the term “gray-box modeling”. The need of these models arises from the limitations of the other modeling approaches such as: white-box can be applied only in exceptional cases, black-box does not give back the right amount of knowledge to the scientist that is using it. Therefore, Design of Experiments (DoE) techniques are often a useful remedy to these issues. In particular DoE techniques can be used to *explain* the behavior of systems treated with black-box models that can only *describe* the behavior, or they can be coupled to computer-based experiments (e.g. finite element method CAEs) that can simulate and solve the partial differential equations derived from a physical model. Considering the latter case, very often the coupled systems of partial differential equations make it difficult to understand the relationships between inputs and response variables thus the physical model must be implemented on CAE software. Thanks to numerical solvers, these software are able to return a clearer vision of the output effects. When scientists are approaching something completely new or when complex systems are described by a large number of input variables, DoE tools offer procedures to retrieve the free parameters of the model or “Screening”– in the first case - and the important parameters identification or “Sensitivity Analysis” for the second case [8]. The advantages of the DoE are [8]:

- It is an Objective and Systematic procedure: it offers a clear path to follow for planning and evaluation of the results.
- Offers the possibility to predict and optimize products and process characteristics.
- Offers Screening solutions easy to interpret (Factorial Designs).
- Helps Modeling by acquiring knowledge from cause-effect relationships.
- Helps to reduce the timings during the project lifetime (e.g. during the design, testing and quality controls).
- Reduces the costs of the experiments since it limits the required number of samples needed to extrapolate all the parameters interactions (D-Optimal Designs).
- Can deal with variance-affected parameters.

Some limitations of DoE techniques [8]:

- Require previous knowledge and solid expertise to comprehend the results that are not always obvious.
- The results are linked to the parameter ranges chosen for the experimental design therefore the experience of the experts is fundamental to find useful experimental outcomes.
- The extrapolated results can be nonphysical so they must be compared with white or black-box models.
- The costs of the required experiments can be still excessive due to the different samples required or the time required to carry out the experiments.
- The results must be interpreted via dedicated software (e.g. *CornerStone*, *MiniTab*, *ETAS ASCMO*) in order to be easily read via graphic tools.

An example of the possible DoE designs is shown in the *Figure 7*:

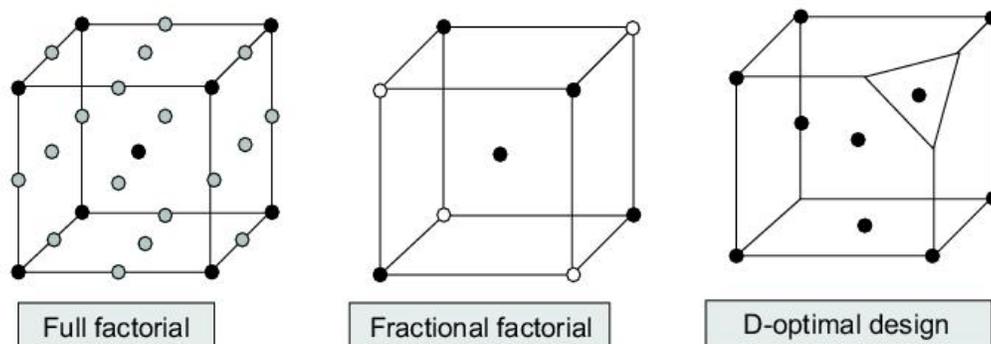


Figure 7 - Examples of DoE designs [15]

3. SOFC PHYSICAL MODEL

In this chapter will be analyzed the white-box model used to describe the behavior of the SOFC's single cell developed by *Bosch* and *Ceres Power*. Understanding the physical laws governing the behavior of the cell is fundamental for the extraction of the interesting parameters and for interpreting the results. The aim is to describe in detail the model by retrieving all the characteristic equations, parameters, factors involved and give a first overlook on how they affect the cell behavior.

3.1 CELL VOLTAGE

Considering the fuel cell as a device that can be powered by different types of fuel, such as hydrogen or hydrocarbons, it's important to define the generalized equation that governs the electrochemical aspects before analyzing in detail their difference. The main concept is the definition of the maximum voltage, also called Reversible Voltage or *Nernst Potential*, that can be generated from an electrochemical reaction. In *Equation 1*, is possible to define it as [16]:

$$E_{Rev}(T) = E^0 + \Delta E(T) = \frac{\Delta G^0}{nF} + \frac{\Delta G}{nF} = \frac{\Delta G^0}{nF} + \frac{RT}{nF} \ln\left(\frac{P_{in}}{P_{out}}\right) \quad (1)$$

Where:

- E_{Rev} is the Nernst reversible potential, i.e. the maximum voltage generated from an ideal reversible reaction ($dS = 0$, with S = Entropy).
- E^0 is the characteristic potential of the reaction generated in standard conditions (i.e. $T = 298,15$ K and $P = 1$ atm), this value is known and can be retrieved from tables.
- ΔE is the difference in potential generated by the reaction according to the absolute temperature at which the reaction is taking place.
- $\Delta G^0, \Delta G$ is respectively the Free Gibbs Energy at standard conditions and with temperature dependence. Both describe reactions that take place at constant temperature and pressure, such as in the case of the SOFC.
- T is the Absolute Temperature and is expressed in Kelvin [K].
- R is the Universal Gas Constant and is equal to $8,314$ [J/mol*K].

- n represents the number of electrons that are exchanged in the electrochemical reaction.
- F is the Faraday's Constant and is equal to 96487 [C/mol].
- P_{in} is the partial pressure of the species of the reactants.
- P_{out} is the partial pressure of the species of the products.

In the following *Table 2* are defined the two reaction that can take place in a SOFC, depending on the type of fueling, and their relative Nernst equation of the electrochemical potentials.

Fuel cell Reaction	Nernst Equation	E_{Rev} (293,15 K, 1 atm)
$H_2 + \frac{1}{2}O_2 \rightarrow H_2O$ $n = 2$	$E_{Rev} = E_0 + \frac{RT}{2F} \ln \left(\frac{P_{H_2} \cdot P_{O_2}^{\frac{1}{2}}}{P_{H_2O}} \right)$	1,229 V (H ₂ O liquid) 1,180 V (H ₂ O gas)
$CH_4 + 2O_2 \rightarrow 2H_2O + CO_2$ $n = 8$	$E_{Rev} = E_0 + \frac{RT}{8F} \ln \left(\frac{P_{CH_4} \cdot P_{O_2}^2}{P_{H_2O}^2 \cdot P_{CO_2}} \right)$	1,060 V (H ₂ O liquid)

Table 2 - Reactions and Nernst Potential of SOFC electrochemical reactions [16]

Reading the Nernst equation in terms of Gibbs' Free Energy is possible to understand the maximum potential of the reaction with respect to the temperature. Indeed, Gibbs' free energy decreases as the temperature increases resulting into a lower cell voltage. This phenomenon is more visible for high temperature fuel cells and SOFCs where, due to the high temperature, water as byproduct is always in the gaseous form resulting into an H₂ oxidation potential always lower than 1,229 V [16].

Although the equations are well defined for each type of reaction taking place in the cell, the cell voltage cannot be described only by its reversible potential. Indeed, fuel cell modeling is based on the approximation of the current-voltage curve $E = f(i)$ (also called "polarization curve") derived from experimental data [3]. Since the curve approximates the real cell voltage behavior all the sources of loss must be considered. The first irreversibility is the presence of an internal resistance of the components and leakages that absorb part of the current, therefore lowering the available generated voltage. The latter can be defined as Open Circuit Voltage E_{OCV} . Starting from the open circuit voltage the curve shows a non-linear behavior due to the presence of three main

sources of losses: Activation, Ohmic and Concentration losses. These losses are always present with different magnitudes along the whole polarization curve although each of them is more dominant in a specific region. Equation 2 is representative of the SOFC voltage and in Figure 8 is presented the current-voltage curve.

$$E_{Cell} = E_{Rev} - \eta_{Losses} = E_{OCV} - \eta_{Activation} - \eta_{Ohmic} - \eta_{Concentration} \quad (2)$$

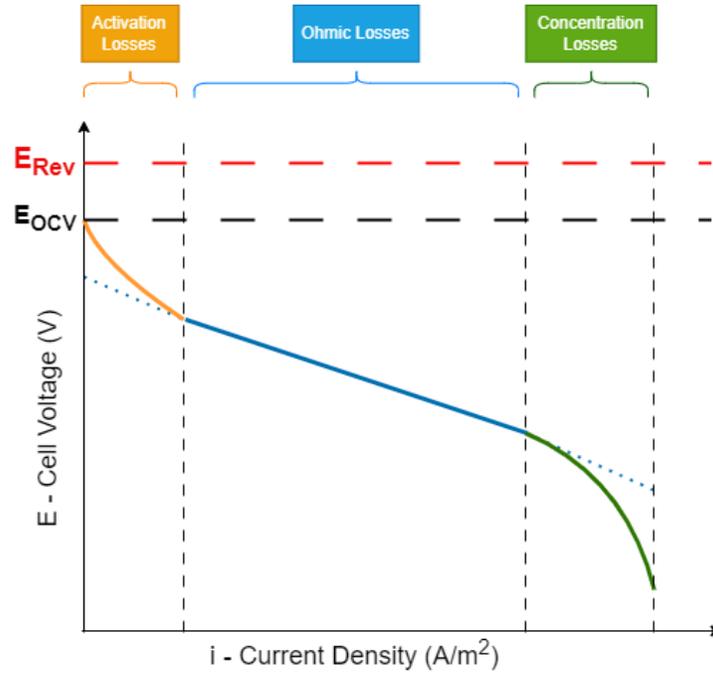


Figure 8 - Polarization curve of a SOFC

3.2 ACTIVATION LOSSES

The activation losses are dominant in the left part of the polarization curve. Their behavior can be described by the *Butler-Volmer* formulation presented in Equation 3 [17].

$$i = i_{0,el} \left(e^{-\frac{\alpha_{el} n F}{RT} \eta_{Act,el}} - e^{\frac{(1-\alpha_{el}) n F}{RT} \eta_{Act,el}} \right) \quad (3)$$

Where:

- $\alpha_{electrode}$ is the *transfer coefficient* that must be known for both electrodes.

- $i_{0,electrode}$ is the electrode *exchange-current density* which must be determined for both the electrodes through the formulas:

$$i_{0,anode} = (\gamma_{0,an}T) \cdot \left(\frac{pH_2}{P_{an}}\right)^a \cdot \left(\frac{pH_2O}{P_{an}}\right)^b e^{\left(-\frac{E_{act,an}}{RT}\right)}$$

$$i_{0,cathode} = (\gamma_{0,cat}T) \cdot \left(\frac{pO_2}{P_{cat}}\right)^c e^{\left(-\frac{E_{act,cat}}{RT}\right)}$$

Considering:

1. $\gamma_{0,an}, \gamma_{0,cat}, a, b, c$ are values specific for the electrode and derived from experimental analyses.
2. P_{an}, P_{cat} is the total pressure in each of the electrodes.
3. pH_2, pH_2O, pO_2 are the partial pressure of the species on the interfaces.
4. $E_{act,an}, E_{act,cat}$ are the activation energies typical of the material.

The complexity of the function renders it impossible to retrieve a direct formulation for $\eta_{Activation}$. Therefore two possible paths can be followed:

- Retrieve the value from the solution of the ODEs at the triple phase boundaries (i.e. the interface between the electrode and the electrolyte).
- Linearize the equation in the *Tafel* form [3]:

$$\eta_{Act,el} = \frac{RT}{\alpha_{el}nF} \ln\left(\frac{i_0}{i}\right) \quad (4)$$

The second one only approximates a certain portion of the activation losses $\eta_{Act} < 0.01 V$. The first one is more precise but more computational expensive, due to the nature of the losses this procedure was used.

3.3 OHMIC LOSSES

Ohmic Loss occurs for two main reasons: electrical resistance and conductivity resistance of the materials. The electrical resistance can be approximated to the thickness of the components δ , while their ionic conductivity σ must be calculated basing on experimental data. Ohmic Losses show a linear behavior and can be

represented by *Equation 5*. They are dominant in the middle range of the polarization curve [3].

$$\eta_{Ohm} = R_k \cdot i = \sum_k^{N \text{ layers}} dV_k \quad (5)$$

Where:

- R_k is the so-called Mixed Conductivity Resistance of each component and is equal to $R_k = \delta/\sigma$ [m²/S].
- δ is the thickness of the layer [m].
- σ is the ionic conductivity and is dependent on the temperature and the material used, e.g. $\sigma(T) = \sigma_0 \cdot e^{-\frac{E_{act}}{RT}}$ [S/m] for the Solid Oxide electrolyte.

3.4 CONCENTRATION LOSSES

In the last part of the curve, the Concentration losses (or Diffusion losses) are dominant. These losses depend on the type of reaction and on the current drawn from the cell. In particular, the latter occur due to the transport of gases in a perpendicular direction with respect to the surface of the electrodes. The flow is described by diffusion laws since the electrodes are made of porous material, i.e. the path to be traveled is a duct with irregular shape and a very small diameter. Due to the complexity of the phenomenon, different considerations and experimental evaluations are needed before choosing the correct model available in literature. Hereafter different diffusion models with their respective applicability are presented [3]:

- **Fick's Laws:** describe molecular diffusion through pores with diameter larger than the average length of the path ($d \gg l$). These laws connect the transport of mass to the concentration gradient with the mass balance inside the layer. The objective is to retrieve a "D" diffusion coefficient that approximates the experimental data. Fast and easy to implement.
- **Stefan-Maxwell equations:** describe diffusion in multicomponent systems for diluted gas and liquids. These equations are quite complex and require different coefficients to be determined a priori.
- **Dusty Gas Model (Knudsen Diffusion):** this model is implemented when the radial size of the pores is comparable to the path to be traveled by the molecule.

Due to the nature of the thicknesses and the porosity of the electrodes in the SOFC, this model was employed in order to describe the diffusion processes in the physical model [17]. In the *Equation* the diffusion coefficient D is presented:

$$D = \frac{\lambda_{path}}{3} \sqrt{\frac{8RT}{\pi M_i}} \quad (6)$$

Where:

1. λ_{path} : is an empirical parameter that is linked to the *diameter of the pores* (d), the *porosity* of the material (defined as percentage of the void space with respect to the volume), the *tortuosity* τ of the duct (defined from experimental data and describes the irregularity of the path and its friction).
2. R : is the universal gas constant.
3. T : is the absolute temperature expressed in Kelvins.
4. M_i : is the molecular mass of the species in exam.

Starting from the Knudsen Diffusion model, different empirical models and equations can be used to determine the diffusion losses. In *Equation 7* is presented the equation used to determine the overpotential resulting from diffusion and mass transport [17]:

$$\eta_{conc} = \frac{RT}{2F} \ln \left(\frac{p_{H_2O_{TPB}} \cdot p_{H_2}}{p_{H_2O} \cdot p_{H_2_{TPB}}} \right) + \frac{RT}{4F} \ln \left(\frac{p_{O_2}}{p_{O_{2TPB}}} \right) \quad (7)$$

Where the subscript TPB identifies the partial pressures of the species at the Triple Phase Boundary. The equation can be read from left to right as the summation of the anode and cathode concentration losses respectively. The partial pressures of the species at the TPB can be written as function of the current density i , as [17]:

- $p_{H_{2TPB}} = p_{H_2} - \frac{RT \cdot \tau \cdot i}{2DF}$: D as diffusion coefficient and τ as the tortuosity both for the anode.
- $p_{H_2O_{TPB}} = p_{H_2O} + \frac{RT \cdot \tau \cdot i}{2DF}$: D as diffusion coefficient and τ as the tortuosity both for the anode.

- $pO_{2_{TPB}} = P - (P - pO_2) e^{\left(\frac{RT \cdot \tau \cdot i}{4DFP}\right)}$: with P as Total Pressure and D as diffusion coefficient both for the cathode, τ as the tortuosity of the cathode pores.

3.5 DISCUSSION ON THE PHYSICAL MODEL

Due to the very complex multi-physical nature of the model, it is clear the need to adopt different modeling solutions that don't require solving systems of PDEs or ODEs every iteration. Another problem arises from the non-homogeneous operating conditions, such as temperature, molar fraction of hydrogen and current density. This issue force to apply the model multiple times for different points of the cell surface in the form $X_i(T, i, pH_2)$. Therefore, the model was implemented in *COMSOL MultiPhysics* in order to simulate the cell in its entirety. In *Figure 9* is presented a schematic view of the average operating conditions with respect to the cell length coordinate "X":

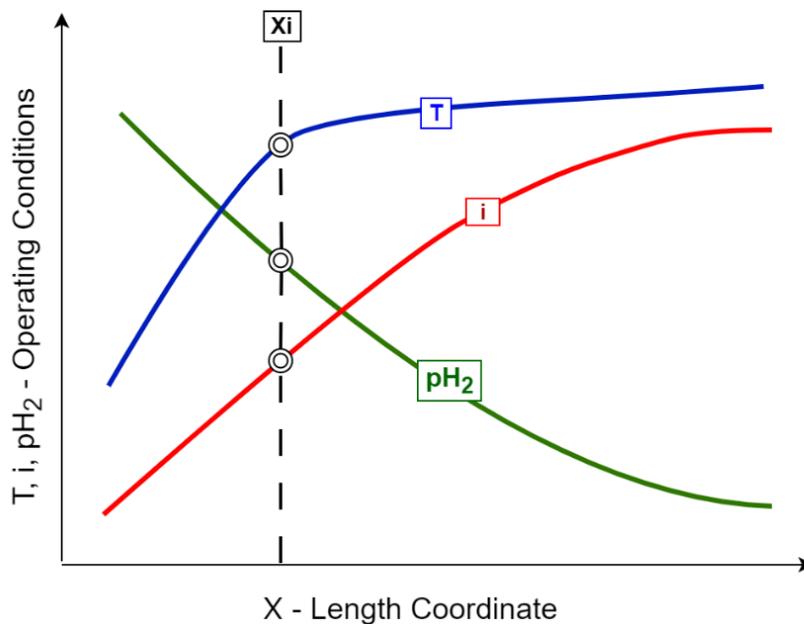


Figure 9 - Average operating conditions across the cell length

Although was possible to connect the model to a *FEM software*, engineers had to face the limits imposed by the time and the model validation procedures. The latter implies the simulation of a very limited geometry that could limit the time expenditure of the

computation, while guaranteeing a limited error on the output. In the following chapters the multi-physics model, exploited in the next steps, will be described in detail with a particular focus on the governing parameters and boundary conditions applied. Nevertheless, the knowledge of physical model is fundamental to retrieve the important actors for the cell voltage and establish the correctness of the results obtained from simulations or other modeling solutions. Some considerations can be made starting from the physical model:

- Geometrical parameters, such as thicknesses or radial dimensions, play a key role in affecting the voltage losses. As an example, ohmic losses are directly linked to the thicknesses of the layers in analysis.
- The activation losses are governed by factors linked to the material properties such as activation energies, pre-exponential factors and operating conditions like Temperature.
- Diffusion losses depend upon the porous materials characteristics such as thickness, porosity, tortuosity of the pores and operating conditions.

4. SOFC SIMULATION MODEL

In this chapter will be analyzed the *COMSOL MultiPhysics* model used to evaluate the cell voltage and the losses breakdown. As discussed before, this model was built to evaluate these properties considering all the design features, the material properties and the real operating conditions. Due to the complexity of the phenomenon and the relative long computational time required, the scale of this model represents a very small portion of the cell.

4.1 GEOMETRY

The model's geometry represents the four main components of the single repetitive unit: Anode's substrate, Anode electrode, Solid Oxide Electrolyte and Cathode electrode. The substrate is a particular characteristic of this product and fulfils different roles: it supports the three active layers, it's responsible for the cooldown of the cell thanks to its high thermal conductivity and it delivers the fuel to the anode through laser drilled holes. The geometry was designed to exploit axial rotational symmetry which is able to reduce the computational load of the simulations. The radial volume analyzed is chosen as the half of the laser drilled holes distance d to maximize the coverage of the active surface of the cell. In *Figure 10* is presented the model geometry:

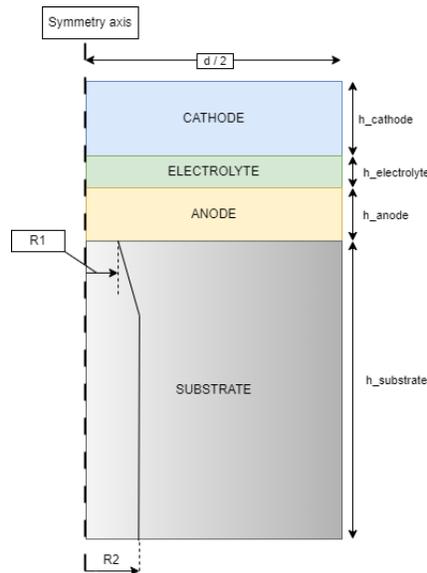


Figure 10 - Simulation geometry

Where:

- " $h_$ " represents the thicknesses of the different components.
- $R1$ is the radius of the laser drilled hole at the Substrate outlet.
- $R2$ is the radius at the inlet of the substrate, dependent on $R1$.
- d is distance between two drilled holes.

4.2 PHYSICS AND BOUNDARY CONDITIONS

Setting the boundary conditions is a fundamental step in the *FEM* evaluations. This model requires the definitions of different physics that can be set by the following *COMSOL* features: *Chemistry of species*, *Electric Current* and *Transport of Concentrated Species (TCS)*. The first feature is used to set the gas species characteristics according to their chemical properties, the second one determines the exchanged current direction and magnitude across the surfaces. The latter defines the transport of the gas species according to the diffusion laws and the mass conservation equation. Since every of these features requires specific boundary conditions, the latter are treated for every component in the next subchapters.

4.2.1 CATHODE DOMAIN

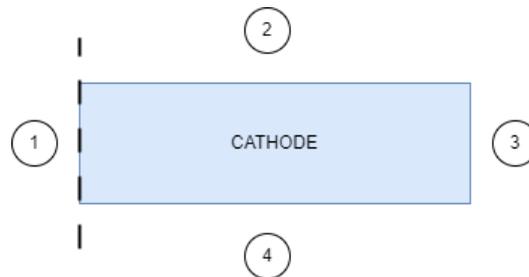


Figure 11 - Cathode's Domain

Considering that the Cathode is a porous layer, Transport of Concentrated Species describes the mass transport according to the imposed Knudsen Diffusion model. The equation solved by this feature over the domain is:

$$\nabla \cdot \mathbf{J}_i + \rho(\mathbf{u} \cdot \nabla)\omega_i = R_i \quad (8)$$

i.e. the Continuity or Transport Equation in case of a Convection-Diffusion problem in Steady State conditions, where:

- $\nabla \cdot J_i$ is the Diffusion Term and it's defined with the divergence of the Diffusion Flux J_i , specific of the species. The Diffusion Flux J_i is linked to the Knudsen diffusion model via the diffusion coefficient D previously discussed in Equation (6).
- $\rho(u \cdot \nabla)\omega_i$ is the Convection Term composed by the average density of the gas ρ , $(u \cdot \nabla)\omega_i$ is the gradient of the velocity field u and the mass fraction ω_i .
- R_i is the Source Term linked to the Cathode's Reduction reaction.

Regarding the Electric Currents module also the Continuity Equation is imposed as:

$$\nabla \cdot J = Q_j \quad (9)$$

Where:

- $\nabla \cdot J$ is the divergence of the Current density flux where J is the generalized Ohm's law in the form: $J = \sigma E + J_e$, where σ is the Electrical conductivity [S/m^2], E is the electric potential [V] and J_e is an external source of current.
- Q_j represents the Source Term.

Taking *Figure 11* as a reference for indexes, the boundary conditions applied to the domain are:

1. *Axial Symmetry* condition was imposed for all the physics.
2. *Initial conditions* such as Pressure (P), Temperature (T) and Air Mass Flow Rate (\dot{m}_{air}) were imposed for the TCS physics, while a *Floating Potential* was defined in order to retrieve the relative electrical potential from equations.
3. *No Flux* condition and *Electrical Insulation* were imposed on this boundary.
4. *Mass Flux* at the interface was imposed equal to the equivalent mass of the Oxygen ions that reacted in the cathode taken as the result of the diffusion in the material. For the electrical side, the Voltage was set as the electrolyte potential subtracted from the *ODE's* result.

4.2.2 ELECTROLYTE DOMAIN

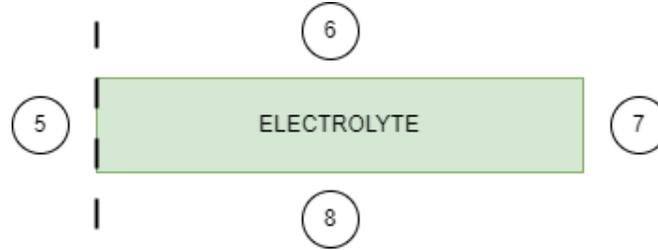


Figure 12 – Electrolyte's Domain

Considering the Electrolyte domain in *Figure 12*, on boundaries 6 and 8 were imposed two *Ordinary Differential Equations* that are responsible for the evaluation of each of the electrode's overpotential at each triple phase boundary. The latter were imposed as:

$$f_{TPB} = C \frac{(dEdl)_{TPB}}{dt}$$

Where:

- f_{TPB} is the Source Term represented by the Current Density difference in boundary [A/m^2].
- C is the specific double layer Capacitance [F/m^2].
- $dEdl$ is the Voltage generated over the boundary [V].

Since the analysis was carried in Steady State, the last term of the equation is null thus reducing the equation to an equivalence between the two currents at the interface. Considering the indexing in *Figure 12*:

1. *Axial Symmetry* condition was applied.
2. In the TCS module the direction of the current density was imposed towards the cathode. The *boundary ODE* was imposed at the triple boundary interface.
3. *No Flux* conditions imposed for both physics.
4. *Boundary ODE* was set to account for the triple boundary interface between electrolyte and anode. Its result was used to set the voltage at the interface.

4.2.3 ANODE DOMAIN

For this domain, the same equations as for the Cathode's Domain were solved thus Convection-Diffusion Equation (8) for the TCS module and the Continuity Equation (9) for the Electric Current Module.

The boundary conditions imposed are referred to the indexing used in *Figure 13*:



Figure 13 – Anode's Domain

1. *Axial Symmetry* condition applied.
2. *Mass flow rate of the species* was imposed as the result of the anode's reaction for the TCS feature. The normal direction of the current was imposed in the electric module.
3. *No flux* condition and *Electrical Insulation* were applied on this boundary.
4. *Fuel mass flow rate* (\dot{m}_{fuel}) was imposed for the TCS and *Electrical Insulation* for the current.

4.2.4 SUBSTRATE DOMAIN

Considering the indexes in *Figure 14*:

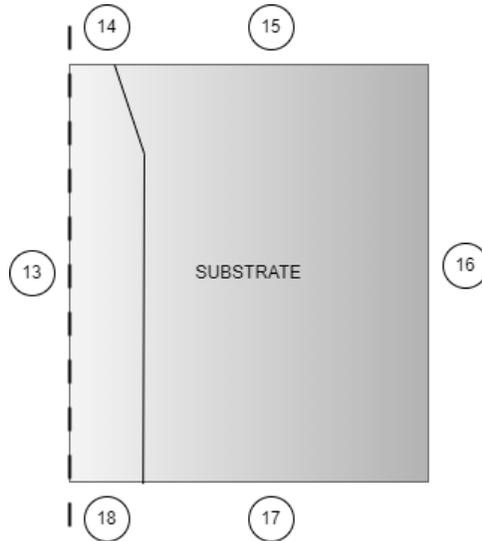


Figure 14 – Substrate's Domain

1. *Axial Symmetry* condition was applied.
2. Fuel mass flow rate (\dot{m}_{fuel}) was imposed for the TCS and *Electrical Insulation* for the current module.
3. *No Flux* condition was applied for the TCS feature.
4. *No Flux* and *Electrical Insulation* were imposed on this boundary.
5. *Ground* condition was imposed ($V = 0$).
6. Were imposed the *Inlet initial conditions* and the concentration of the fuel species χ_{fuel} .

4.3 WORKFLOW AND EXPECTED RESULTS

The purpose of this steady-state model is to evaluate the cell behavior in different scenarios. Due to the nature of the equations presented in the physical model section and the boundary conditions previously discussed, we need to make a distinction according to the type of fuel provided to the cell, in particular:

- **Humidified Hydrogen:** the model requires a manual setting of the molar fraction χ of the fuel species, in this case only H_2 and H_2O . The current density i is set in order to simulate the absorbed by the load connected to the SOFC

stack. The temperature T and the molar fraction determine which part of the cell is being simulated.

- **Hydrocarbon based (WGSR):** the species composition of the fuel is linked to the Fuel Utilization Factor FU which determines the real mass flow rate of fuel that reach the anode side. T and i are also varied accordingly to the purpose of the investigation.

The workflow followed by the software is hereby presented:

- i. The model is set according to the fuel that is being studied. The input operating conditions are set according to the position coordinate to be investigated.
- ii. The mass flow rate of the fuel and air species is set for both the electrodes' inlet boundaries.
- iii. Ordinary Differential Equations are solved at the two triple phase boundaries of the electrodes and the results determine the potential for each layer interface.
- iv. Activation Losses are calculated for each current density value for both the electrodes with Equation (3).
- v. Ohmic Losses are evaluated according as the sum of the over-voltages generated in each layer, as in Equation (5).
- vi. Concentration Losses are evaluated basing on the solution of the Transport of Concentrated Species imposed on the electrodes, as Equation (7).
- vii. The cell voltage is evaluated as the difference of the anode and cathode's voltage.

It's important to remark that, due to the small dimension scale of the simulation model and the different operating conditions across the cell, it is necessary to run a large number of simulations to cover completely the cell surface. Therefore, different modeling approaches will be investigated and compared in the following chapters.

5. SENSITIVITY ANALYSIS

Sensitivity Analysis is a decision-making tool which helps the scientists to understand the effects of an independent variable change (input) on a dependent variable (output). This analysis is pursued to investigate new designs as well as the effect of manufacturing tolerances ranges on the current design parameters. The latter is also referred to as Screening Procedure of a model and it aims to identify and select the most impactful parameters from a large pool.

The aim of the next subchapters is to review the Sensitivity Analysis procedure in all its steps, from the preliminary parameter selection to its results.

5.1 PRELIMINARY PARAMETER SELECTION

To get an overlook of the parameters affecting the SOFC voltage behavior is therefore important to make a first selection from all the actors of the multi-physical model. This is done in order to limit the number of simulations (and the relative time expenditure) however keeping the focus on the model understanding and a future design improvement. Therefore, the selection was made according to the following criteria:

- Are investigated only the parameters that can be controlled and varied on further steps. (e.g. Activation Energies were not investigated since cannot be modified except by changing the layer's material. Geometrical design parameters or characteristics were prioritized instead.)
- The parameter's investigation range is chosen according to the respective manufacturing tolerances or statistical scatter.
- In order to get a meaningful comparison between the two types of fuel investigated, six common operating points were selected according to the temperature, hydrogen concentration and current density. For the humidified hydrogen model the hydrogen molar fraction was directly imposed, while for hydrocarbon-based fuel, the FU that guaranteed the same hydrogen molar fraction was retrieved.

Given these assumptions, the investigated parameters and relative scatters $\pm\delta$ from their nominal values are listed in *Table 3*:

PARAMETER NAME	SCATTER δ
Anode thickness (h_anode)	$\pm 17,65\%$ [m]
Cathode thickness (h_cathode)	$\pm 20\%$ [m]
Electrolyte thickness (h_electrolyte)	$\pm 20\%$ [m]
Substrate thickness (h_substrateblock)	$\pm 5\%$ [m]
Radius of the laser drilled holes (r_LDH)	$\pm 8,70\%$ [m]
Anode's porosity (eps_poreanode)	$\pm 21,2\%$ [-]
Anode's pores tortuosity (tau_poreanode)	$\pm 10\%$ [-]
Anode's pore radius (r_poreanode)	$\pm 15\%$ [m]
Cathode's porosity (esp_porecathode)	$\pm 21,2\%$ [-]
Cathode's pores tortuosity (tau_porecathode)	$\pm 10\%$ [-]
Cathode's pore radius (r_porecathode)	$\pm 15\%$ [m]
Geometrical parameter 1 (GP1)	$\pm 20\%$ [m]
Geometrical parameter 2 (GP2)	$\pm 20\%$ [m]
Geometrical parameter 3 (GP3)	$\pm 20\%$ [m]
Material parameter 1 (MP1)	$\pm 21,2\%$ [-]

Table 3 - Parameters and ranges preliminary selection

The geometrical parameters called with the acronym “GPX” represents the thickness of some of the active layers that are active in the single repetitive unite. The material parameter *MP1* at the end of the table represents the porosity of the active layer *GP1*. The information regarding these layers is not shared since the latter are part of a patented design. Regarding the operating conditions, the latter can be summarized in *Table 4*:

Test Case	χ_{H_2}	T [K]	i [A/m ²]	FU
A	0,2884	821,15	1584	0,307
B	0,1802	876,45	2008	0,570
C	0,2600	843,75	1867	0,376
D	0,2767	817,35	1485,7	0,335
E	0,2480	873,65	2202,5	0,405
F	0,1880	891,15	2079	0,551

Table 4 - Test cases for different fuel models

5.2 PROCEDURE

In this thesis work was performed a **One-At-Time (OAT) Sensitivity**, a local-evaluation method, which consists of changing one parameter at the time and then evaluate the output fluctuations accordingly. For this purpose, a code in *Python* was created to run the *COMSOL* simulation models for all the parameters and collect the cell voltage and the losses breakdown as simulations outputs. The simulations were run for six operational points and three geometrical dimensions for each parameter, for a total of 270 runs per fuel model (circa 2 hours).

Once the results were stored, a derivative approach was used in order to represent the output change as a percentage of the input. This approach consists in determining the first derivative of the function in the nominal point of the range, then multiplying it by the overall explored range. The latter can be expressed by the formula:

$$S [\%] = \left. \frac{dy}{dx} \right|_{x_{nom}} \cdot \Delta x \cdot 100 = \frac{y_i - y_{nom}}{|x_i - x_{nom}|} \cdot (x_{max} - x_{min}) \cdot 100$$

Where:

- y_i is the output that is being evaluated (e.g. cell voltage or one of three losses).
- x_i is the input parameter that is currently being evaluated.
- x_{max}, x_{min} are respectively the maximum and the minimum value in the parameter range. (i.e. $x_{max} = x_{nom} + \delta, x_{min} = x_{nom} - \delta$).
- x_{nom}, y_{nom} are respectively the center of the input range and its relative output value.

5.3 RESULTS

To obtain a good overlook over the parameters change influence on the cell voltage, the Voltage Sensitivity was averaged over the six test cases. In this way was possible to obtain only two plots establishing the percentual change in the output, for both the Humidified Hydrogen fuel model and the Hydrocarbon Based fuel model. The results are presented in the so-called “Tornado Plots” in *Figure 15* and *16*:

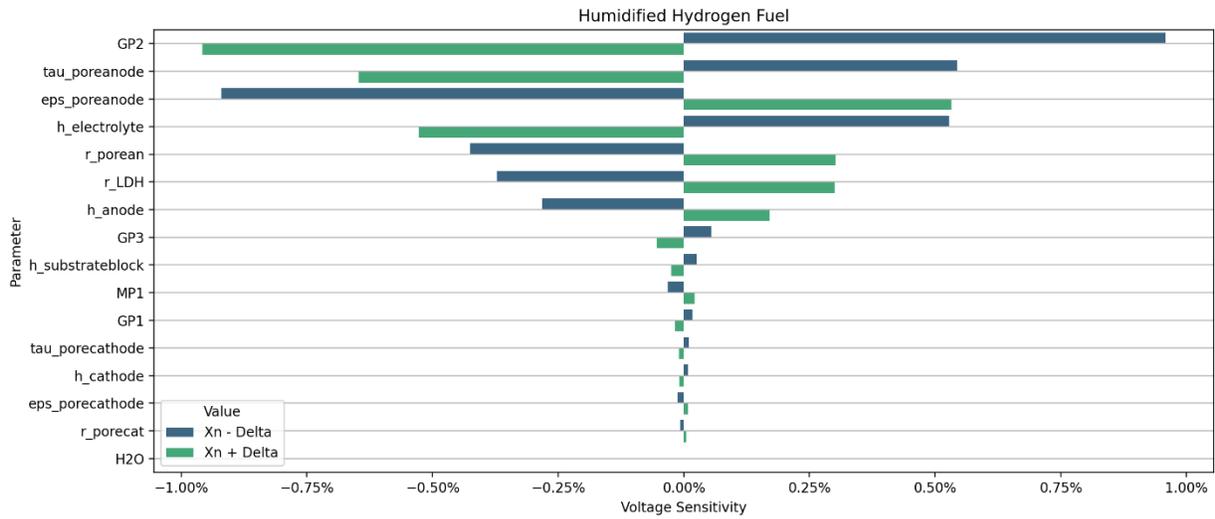


Figure 15 - Parameter Voltage Sensitivity Humidified Hydrogen

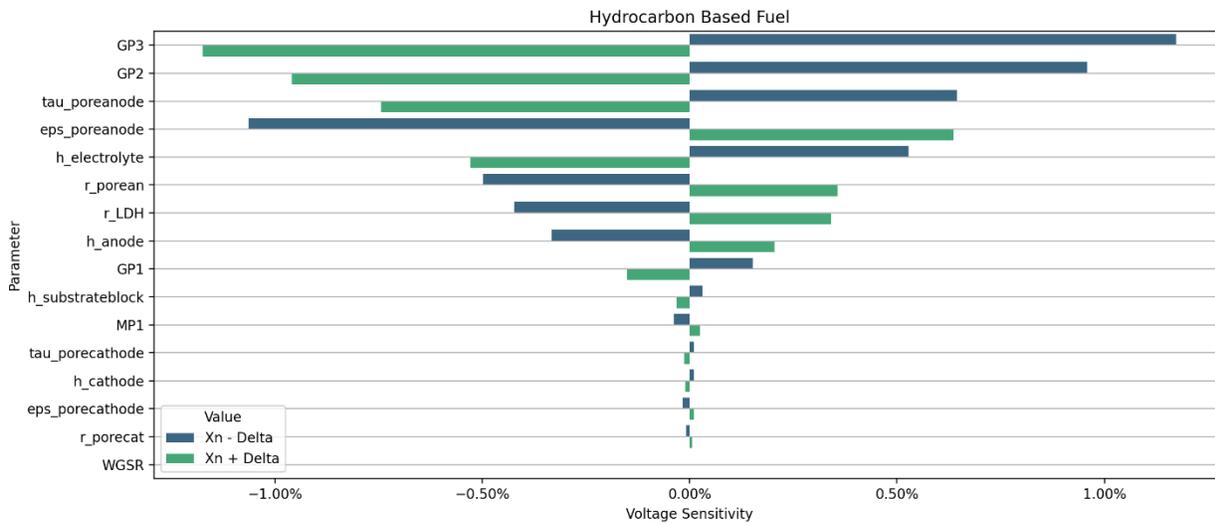


Figure 16 - Parameters Voltage Sensitivity Hydrocarbon Based

From the previous plots is possible to determine which are the parameters that most affect the cell voltage and its relative variation magnitude. One of the major differences denoted from the two types of fuel studied, is the magnitude change for the geometrical parameter identified by *GP3* that appears heavier for hydrocarbon-based fuel model. The latter can be addressed to the nature of the material that characterize the layer which heavily affects the Ohmic Losses in a way directly proportional to its thickness. Another important result is the magnitude difference of the parameters related to the

Anode with respect to the Cathode electrode. This can be addressed to the strong importance that the fuel electrode plays in fuel cells. Indeed, the Cathode most of the times works with an excess of air, therefore never face problems such as lack of oxygen or concentration gradients. The Anode characteristics instead, strongly influences the Concentration losses as well as the Activation losses. For this reason, the Anode is the only geometrical parameter that will induce a loss decrease with an increasing thickness. If we focus on the material parameters such as porosity, tortuosity and radius of the pores, in general we can affirm they all show the same trend connected to the reduction in Concentration losses. By imposing shorter and larger pore paths with lower tortuosity is possible to obtain the best condition to minimize this type of losses. An ulterior search was performed focusing on the losses type and its results are summarized in *Table 5*:

Parameter	Activation Losses	Ohmic Losses	Concentration Losses
Anode thickness (h_anode)	✓		✓
Cathode thickness (h_cathode)	✓		✓
Electrolyte thickness (h_electrolyte)		✓	
Substrate thickness (h_substrateblock)	✓		✓
Radius of the laser drilled holes (r_LDH)	✓	✓	✓
Anode's porosity (eps_poreanode)	✓		✓
Anode's pores tortuosity (tau_poreanode)	✓		✓
Anode's pore radius (r_poreanode)	✓		✓
Cathode's porosity (esp_porecathode)	✓		✓
Cathode's pores tortuosity (tau_porecathode)	✓		✓
Cathode's pore radius (r_porecathode)	✓		✓
Geometrical parameter 1 (GP1)	✓	✓	✓
Geometrical parameter 2 (GP2)		✓	
Geometrical parameter 3 (GP3)		✓	✓
Material parameter 1 (MP1)	✓		✓

Table 5 – Parameter variation influence on the losses

From this analysis it is possible to establish the correctness of the results basing on the physical model discussed before. Anyway, is possible to underline the influence of almost every parameter over the activation losses which equation was not explicitly linked to other parameters than the electrodes. It is important to keep in mind that the results of the previous analysis consider the output changes when only one input parameter is varied every time. Therefore, if the purpose of the analysis is to investigate the model behavior after multiple changes (e.g. design optimization), the scientist needs to refer to other types of modeling, since the interactions cannot be directly evaluated from the results presented above. Indeed, the influence of a multiple parameter change can result, instead, into a different or opposite cell behavior, with a relative different magnitude compared to the more obvious “sum” of the multiple OAT Sensitivities. In the next chapter is presented the Gray-Box analysis with the respective global Sensitivity screening and the full interactions matrix.

Although the results are limited to the single parameter influence the analysis is still very useful for a general model understanding and can be also exploited to lower the number of investigated parameters during the Black-Box training procedure, leading to better accuracies and time expenditure optimization.

Is also essential to remind that the results are linked to the specific SOFC product in analysis via a simulation model which also carries, due to its mathematical nature, a linear error in the results. Therefore, all the results obtained need to be proved by experimental results carried out on a real cell prototype.

6. SOFC GREY-BOX MODEL

In this section will be analyzed the grey-box modeling of the SOFC single cell. As previously mentioned, these techniques allow scientists to explain the model behavior through a synergetic combination of statistical analysis and physical laws. This modeling approach was pursued mainly due to:

- Create a random simulation database with multiple parameters combinations to be used as a training set for the black-box modeling of the cell.
- Having a solid reference for comparison during the black-box testing, since the DoE software is also able to make predictions based on the regression of the provided data.
- Create a first connection between the smaller dimension scale of the simulation model and the larger one represented by the complete cell.
- Derive a new Sensitivity Analysis with higher resolution (due to the larger dataset provided) and compare it with the previous results.

6.1 SIMULATION DATASET

The very first step for a numerical design of experiments is the creation of a large quantity of data. This is because having as much data as possible allows to approximate complex systems with better results and reliability. The latter is also taken as a rule for black-box training procedures.

As previously discussed, the model in analysis deals with 15 different design parameters and 3 different operating conditions (i.e. *Temperature (T)*, *Current density (i)* and *Molar Fraction (χ_{H_2})* for the humidified hydrogen or *Fuel Utilization factor (FU)* for the *WGSR* model).

Each one of these is defined into a specific range delimited with a minimum and a maximum value, thus the output resolution is strictly related to the number of points γ (also called “levels”) selected between the two extremes. If we want to use a full-factorial design to have all the possible combinations of the model parameters on γ levels, we will need 18^γ simulations, each one with a duration between 20 and 30 seconds. Taking $\gamma = 3$ as a minimum, the simulations would require just 48,6 hours, increasing this value to $\gamma = 5$ can be easily determined that the plan would require *just*

2624,4 days (the equivalent of 7,2 years). Therefore, it is unavoidable to apply a fractional-factorial design (also called the *Space-Filling* approach) with a fixed number of simulation points which span on different γ levels over the parameters' ranges. The selection of the simulation points was based on the tradeoff between data and time expenditure, thus choosing $n = 10000$ unique combinations as the data dimension. To have a random dataset with the highest coverage of all the possible cases, the latter was built using *Sobol's Sequence* which is a mathematical series that is also used in cybersecurity for *One-Time-Password (OTP)* generation. This series creates a quasi-random dataset, with a resolution directly proportional to n , combining all the parameters on the selected levels. Therefore guaranteeing the right level of randomness to avoid fake pattern recognition in both the grey and black-box models. In the next *Figure 17* is shown a schematic representation of the combinations treated, for both fuel models, in the three-dimensional space of the operating conditions.

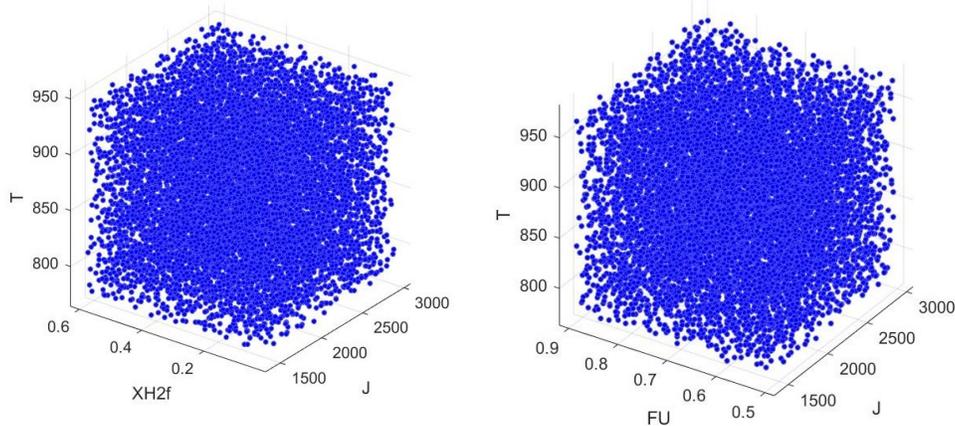


Figure 17 - Schematic view of the Space-Filling plan for both fuel simulation models

Once the dataset was defined, the simulations were run on *COMSOL MultiPhysics* through a code developed in *Python* based on the *Mph* library [18]. The output of the simulations (i.e. Cell voltage and Losses Breakdown) was stored and evaluated through a DoE Analysis software *ETAS-ASCMO* of the *Bosch Group*.

6.2 EVALUATIONS

After collecting the results of 20000 runs in total for both models, the data was treated in the *ETAS-ASCMO* environment. The software provides many different options for

the statistical regression of the input data either with proprietary or Machine Learning based models. Also, other options can be used in the post-processing phase such as the Sensitivity Analysis and the full interaction matrix for multiple parameters. For the sake of clearness, the results will be divided into two different subchapters based on the fuel type simulation model that is being analyzed.

6.2.1 HUMIDIFIED HYDROGEN MODEL

The model was run for different values of hydrogen fraction and complementary water steam concentration for a total of 8996 meaningful runs out. *Figure 18* shows the behavior of the model after the regression made by the default process embedded in *ASCMO (Gaussian Process)*.

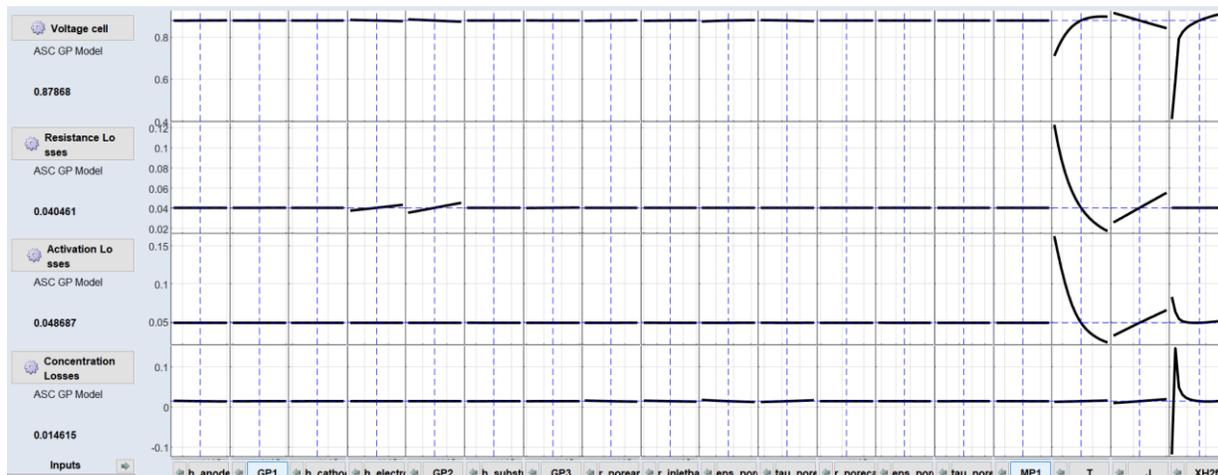


Figure 18 - NDoE results Humidified Hydrogen

In this section of the software is possible to get all the predictions by varying manually all the parameters and reading the relative outcome on the left-hand side of the screen. It is important to remember that the predictions are made only for a pre-defined range and dataset, thus no extrapolation can be made. The model's regression performance can be accessed by the appropriate command and the results are shown in *Figure 19*:

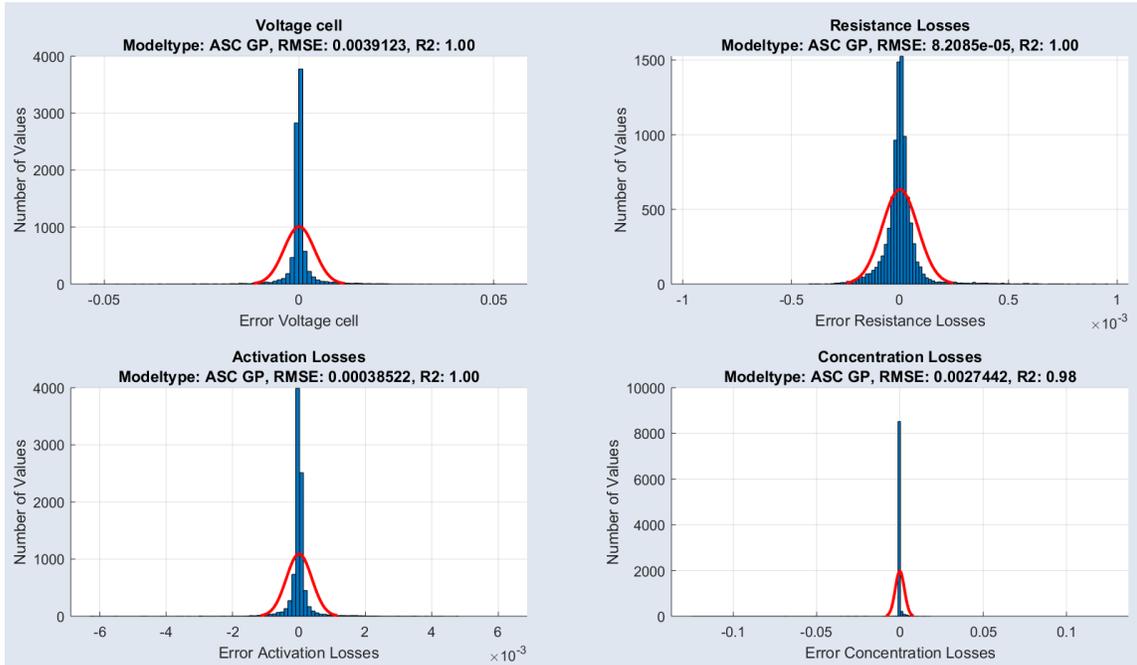


Figure 19 – Prediction Performance GP, Humidified Hydrogen

The good performance of a Regression in general can be evaluated mostly through two parameters:

- *RMSE (Root Mean Squared Error)* that can be evaluated with the formula and establish the mean error of the predictions:

$$RMSE = \sqrt{\frac{\sum_i^N Err_i^2}{N}} = \sqrt{\frac{\sum_i^N (y_{pred} - y_{true})^2}{N}} = \sqrt{MSE}$$

The lower the RMSE the lower the error derived from the prediction.

- *R² (Coefficient of Determination)* that establishes the distance of the residuals from the data interpolated through the regression:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_i^N (y_{pred} - y_{true})^2}{\sum_i^N (y_{pred} - \bar{y}_{true})^2}$$

On the basis of the parameters explained before, we can assess that the model returns a good interpolation of the input data, thus we had a confirmation that the number of points chosen and the randomness of the dataset worked properly. Going back to the cell model analysis, we were able to access the full interaction matrix for multiple parameters. This is a very strong and meaningful graph that shows in a matrix all the possible combinations of two-parameter interactions, nevertheless, the power of this plot is not easily accessible to an unexperienced “eye”. *Figure 20* below shows the interaction matrix plot for the cell voltage only considering the Anode and Cathode parameters:

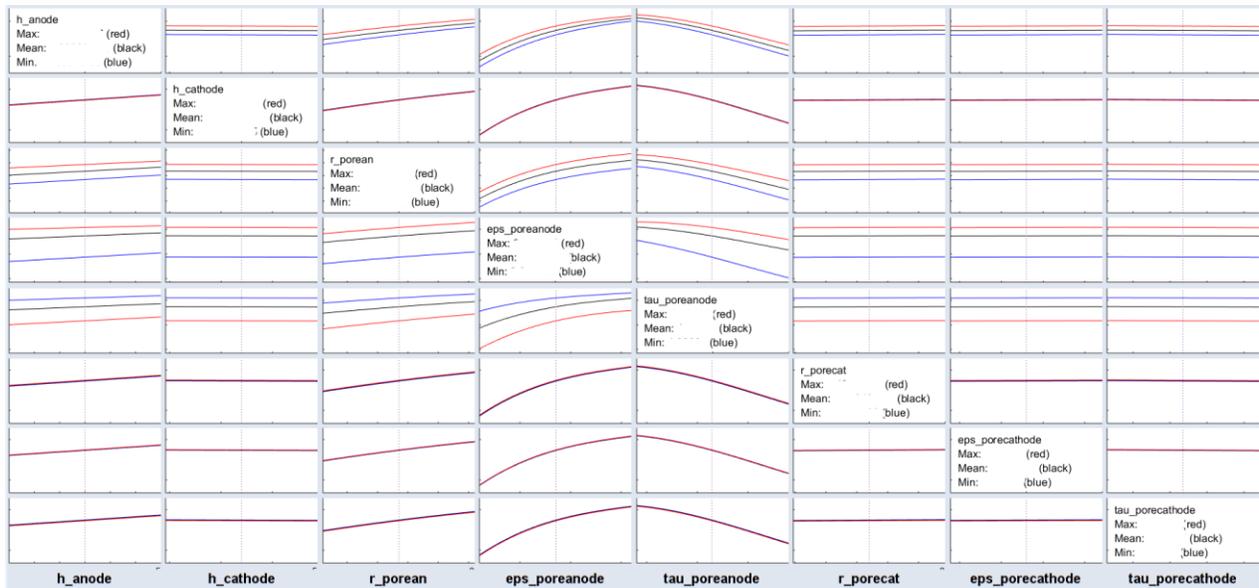


Figure 20 - Interaction matrix for the electrodes related parameters

On the Y-axis is shown the cell voltage, while the three lines represent the extremes and mean value for each parameter on the X-axis. Each *i-th* row displays the cell voltage behavior keeping the *i-th* parameter constant while varying the *j-th* parameter dimension in its relative column. From this plot is possible to assess that the anode’s material parameters such as the radius of the pores, porosity and tortuosity have a major impact on the cell voltage compared to the cathode’s. The second parameter by importance is the anode’s thickness.

These evaluations can be also confirmed by the global Sensitivity Analysis (called *Input Relevance*) derived from the software. As specified in the User Guide, the relevance of the inputs over the output is checked by a stepwise regression where the

inputs with relevance < 5% are ignored. After that, the inputs columns are permuted and a *pseudo-RMSE* is calculated thus getting a heuristic of the input's relevance [19]. The results for this model are shown in *Figure 21* below:

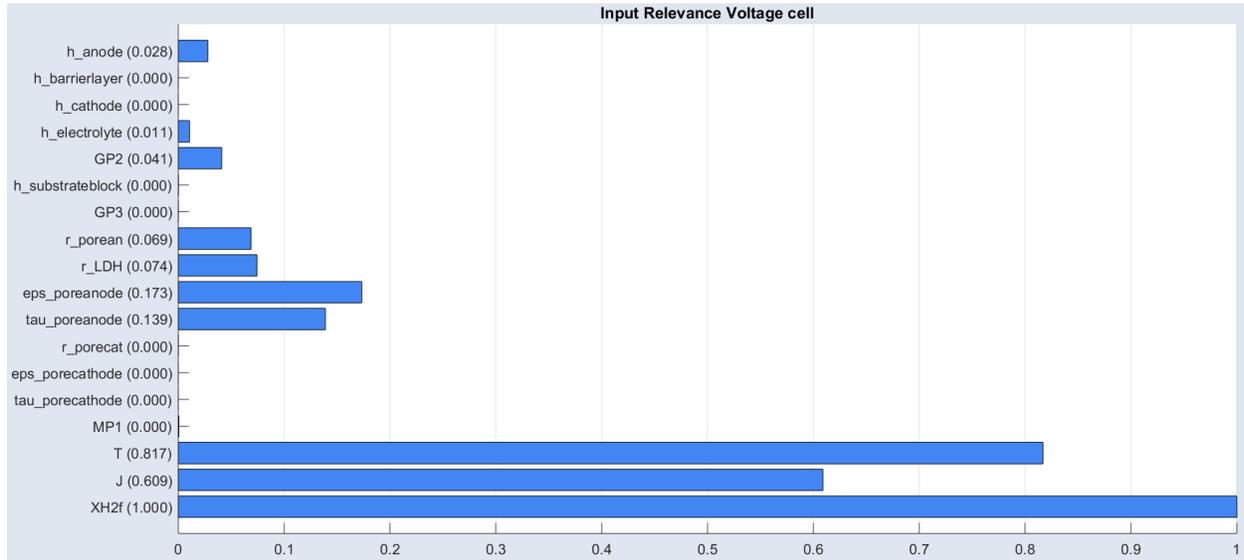


Figure 21 - Sensitivity Analysis Humidified Hydrogen

As it is possible to notice, the three operating conditions have a major role in affecting the cell voltage. This result is somehow obvious if we consider the physical model discussed before where every equation was based on Temperature, Hydrogen Molar fraction and current density. Although excluding these three parameters, the last analysis identifies the same parameters as the previous chapter's but with different weights. Anyway, considering that the software works with a much larger amount of data, the latest results are considered more accurate.

6.2.2 HYDROCARBON BASED MODEL

This model resulted in 8412 positive runs with respect to the 10000 planned in the dataset. The interpolation and relative performance, resulting from the *Gaussian Process* operated by *ASCMO*, are shown in *Figure 22* and *23* below.

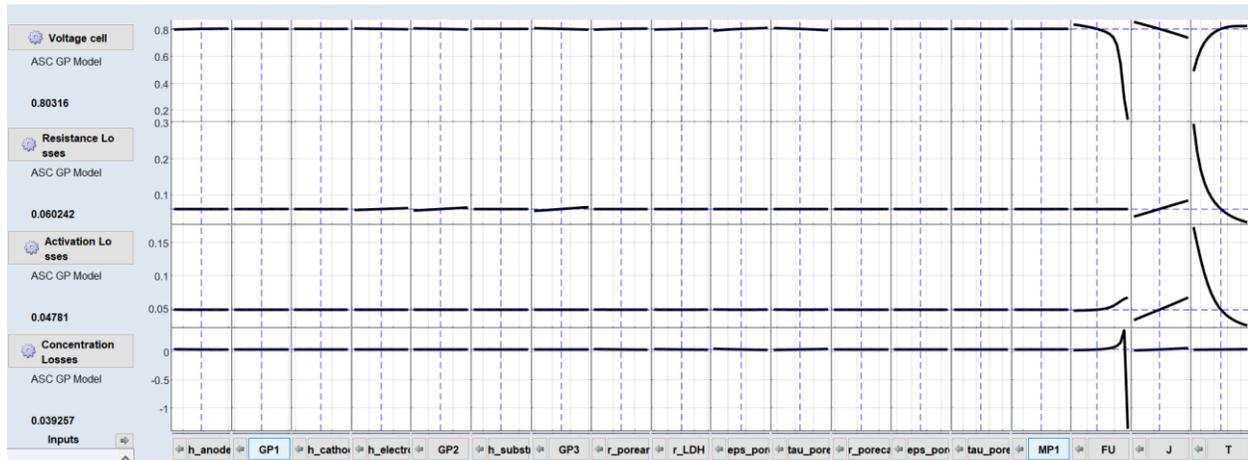


Figure 22 - NDoE Hydrocarbon Based fuel model

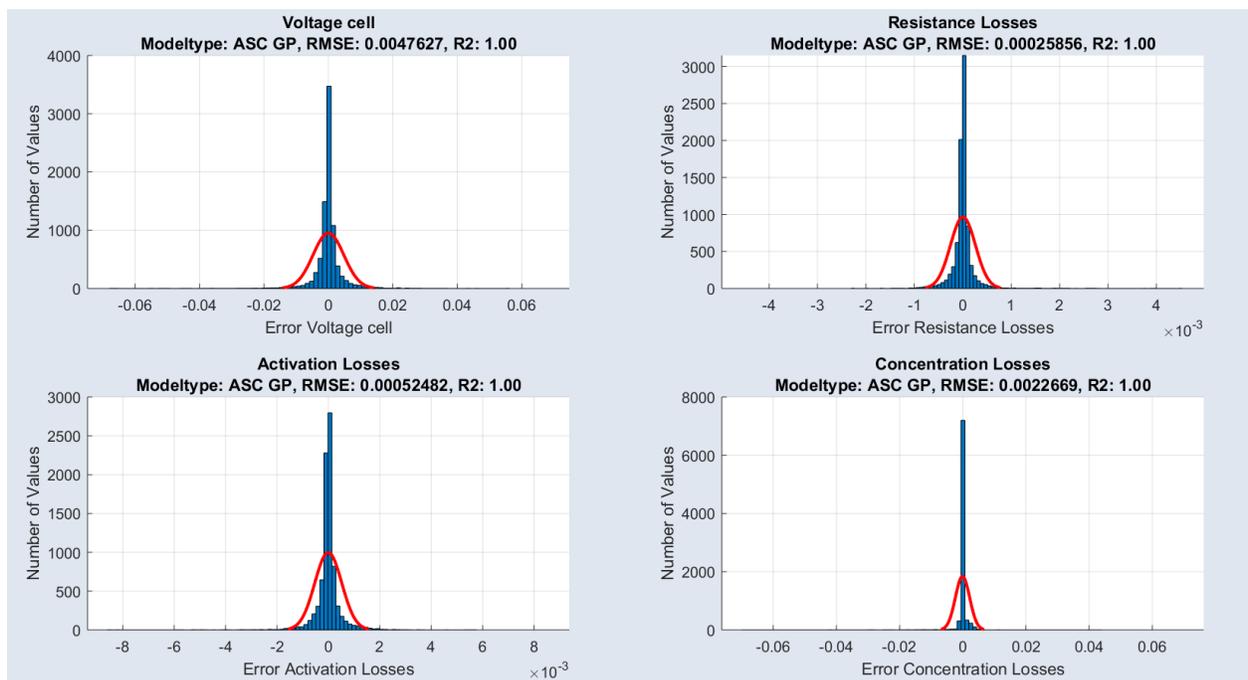


Figure 23 - Prediction performance GP, Hydrocarbon Based

From the pictures above is possible to derive the high precision in prediction operated by the *GP* model. Also, in this case, it was possible to derive the input full interaction matrix for the Anode and the Cathode parameters. The resulting plot is shown below in *Figure 24*.

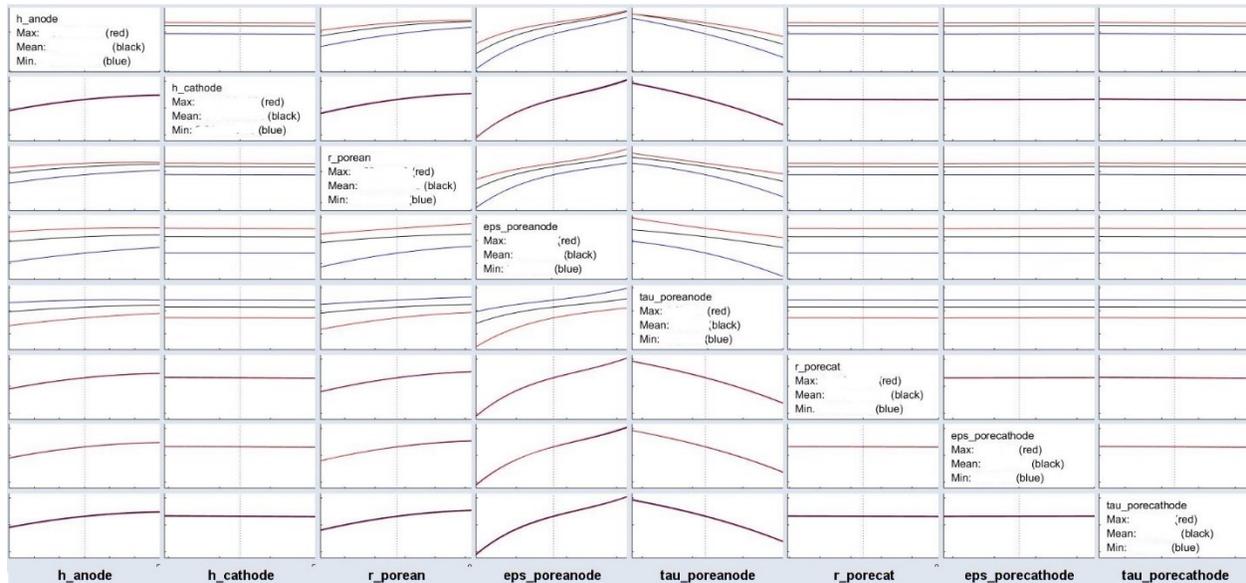


Figure 24 - Interaction matrix for anode's and cathode's parameters

The results are very similar to the humidified hydrogen model, where the anode's material properties are leaders in influencing the cell voltage. Anyway, the behavior is not more approximated by a second-order function (e.g. a Parabola) but it shows a higher-order interaction probably due to the different species present in the fuel composition. The results of the relative Input Relevance are shown below:

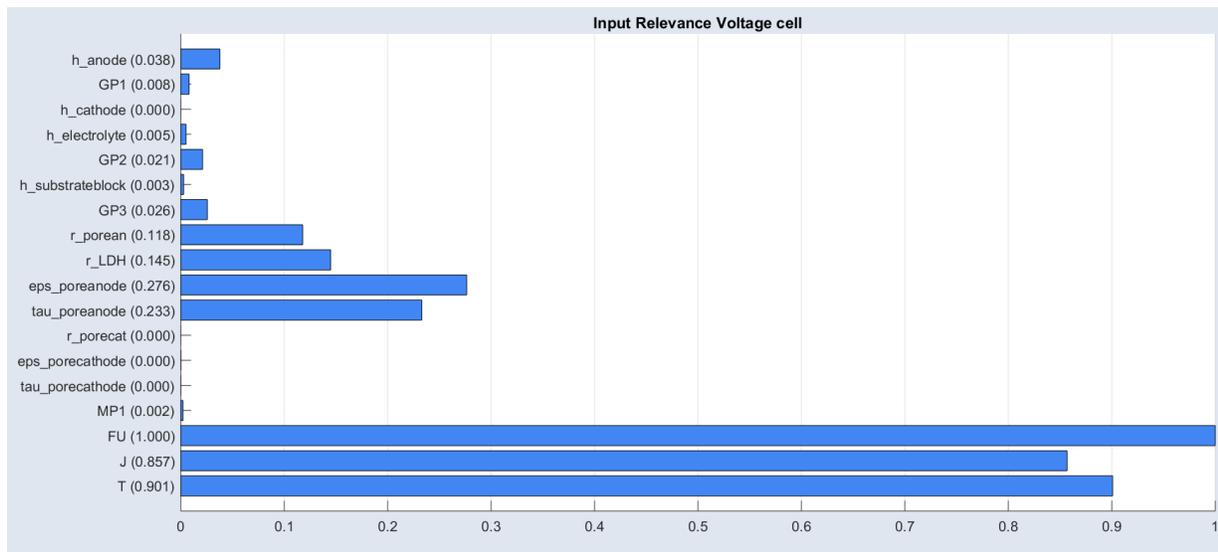


Figure 25 - Sensitivity Analysis Hydrocarbon Based fuel model

For this type of fuel, the relevance of the geometrical parameter $GP3$ is still higher than $GP2$ and the humidified hydrogen model, confirming the previous chapter's results. Indeed, the $GP3$ parameter affects mostly the Ohmic Losses generation of the cell due to its material characteristics and role.

Comparing these results with the humidified hydrogen ones in this section, the most influencing parameters are the same, but they show different importance magnitudes. Anyway, also, in this case, the Anode's related parameters play a key role as well as the radius of the laser-drilled holes (r_{LDH}).

6.3 COMMENTS

Grey-box techniques are very useful tools that can be employed by scientists in different scenarios. First of all, they are used for planning the experiments in the validation phase of a pre-existent physical model, here they provide strong fundamentals for the experiment construction and results evaluation. Another scenario of application is when, due to the lack of prior knowledge or time, grey-box modeling is utilized to approximate and describe the behavior of complex systems while identifying their governing parameters.

In the context of this thesis work, the Design of Experiment was exploited to create a reference for the upcoming steps and to define a random dataset for a Black-box model training and testing. Anyway, without the help of appropriate software, all the information gained and provided in this section most probably would still be hidden in the large amount of data analyzed. Indeed, the use of these software is somehow unavoidable and however, without a good amount of prior knowledge, the results can still be misleading or difficult to grasp.

When working in a collaborative environment made up of different people and independent teams, the second most challenging step is often to present the results of the work done making them accessible to others in a simple and immediate way. Machine Learning tools can be implemented in a coding environment accessible to everyone -and most importantly free of charge- like *Python*. Especially when paired to a custom user interface or embedded into a website, these black-box models can be accessed by everyone with any digital device. Therefore, the mathematical models investigated in the next chapter offer, besides the scientific interest behind this research topic, a practical gain that can be left to the colleagues of the SOFC R&D department.

7. SOFC BLACK-BOX MODELS

Mathematical models offer the possibility to describe complex systems with good accuracies in a matter of a few milliseconds. Concerning this research, a trial-and-error procedure was applied to test different Machine Learning models comparing them in terms of accuracy with the scope of finding the best model that could approximate the SOFC cell behavior. Different models were tested:

- **Linear Models**
- **Classification Algorithms**
- **Artificial Neural Networks**

All the models tested were implemented in *Python* through the library *Scikit-Learn* which offers, in addition to the command to recall algorithms, solutions to avoid common problems such as underfitting and overfitting. Once the best algorithm was found, the latter was analyzed to retrieve its applicability range as it is a fundamental part of the modeling procedure.

7.1 LINEAR MODELS

Linear models are one of the simplest algorithms available for data regression. Their working principle is based on the definition of an approximating function which is responsible for the data interpolation. The algorithms tested that belong to this category are:

- **Linear Regression**
- **Bayesian Ridge Regression**
- **Polynomial Regression**

Even if these models are “basic”, they can predict with good reliability simple models. Nevertheless, their performance degrades increasing the complexity of the analyzed model also due to the lack of *hyperparameters* used to perform finer tuning [13].

7.1.1 LINEAR REGRESSION

Linear Regression algorithm uses a polynomial function of first degree for the prediction of the continuous target function. A polynomial with degree 1 is simply represented by the equation of a straight line, id est:

$$y = mx + b$$

Where:

- y is the continuous target variable.
- x is a discrete input vector.
- m is the slope of the line.
- b is the Y-intercept i.e. the value of the function when the input is null.

In case of multiple inputs (n) the algorithm searches for the best coefficients for the equation:

$$y = m_1x_1 + \dots + m_nx_n + b$$

The objective of the algorithm is the optimization of the slope coefficients m_n and the intercept b to minimize the residuals i.e. the difference between the predicted output y_p and the true output y . This algorithm was implemented in *Python* via the command *LinearRegression()* and the input data was preprocessed through the function *PolynomialFeatures(d=1)*, which determines the shapes of the input vector x based on the degree of the polynomial function used for the regression [11] [12]. The results of the algorithms are presented, for both the fuel models, below in *Figures 26* and *27*:

Model Evaluation Plots for Linear Regression Degree = 1
Cell Voltage - Humidified Hydrogen
 $R^2 = 0.702$ MSE = 0.001742

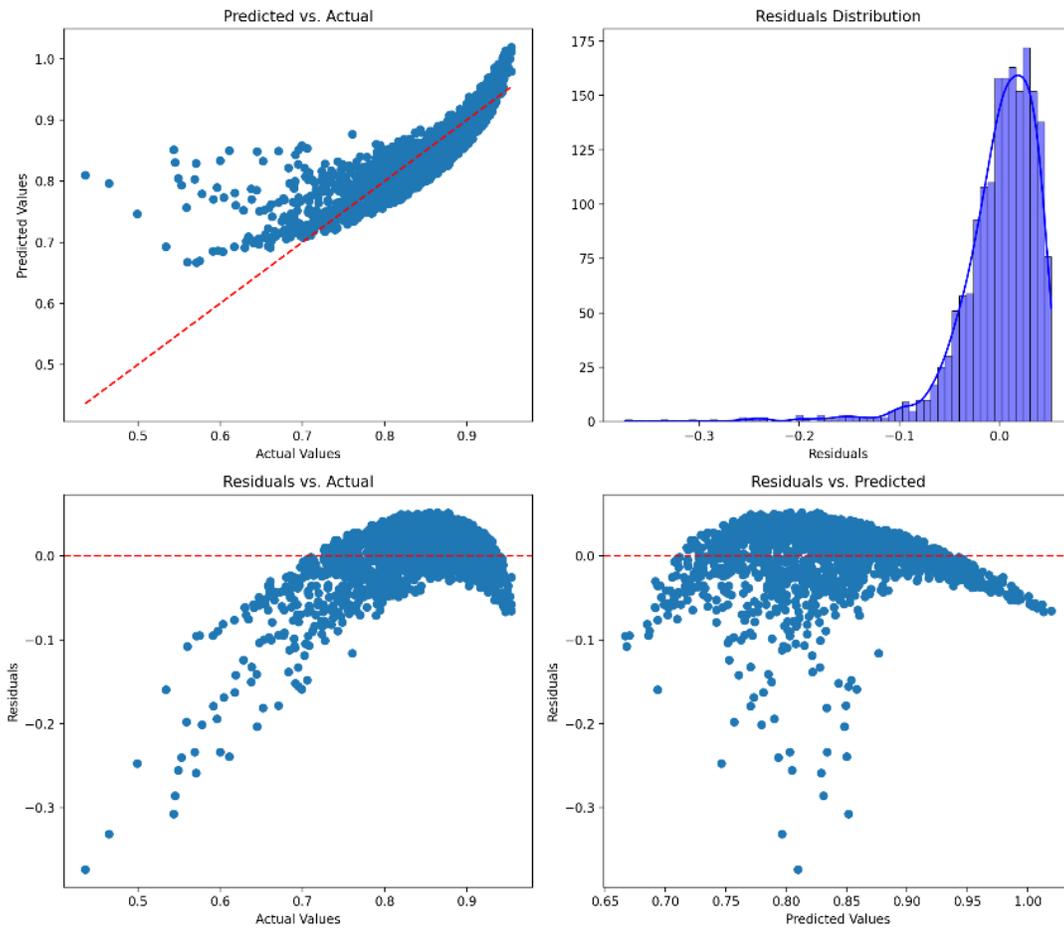


Figure 26 - Linear Regression performance, Humidified Hydrogen

Model Evaluation Plots for Linear Regression Degree = 1
Cell Voltage - Hydrocarbon-based fuel
 $R^2 = 0.7073$ $MSE = 0.003745$

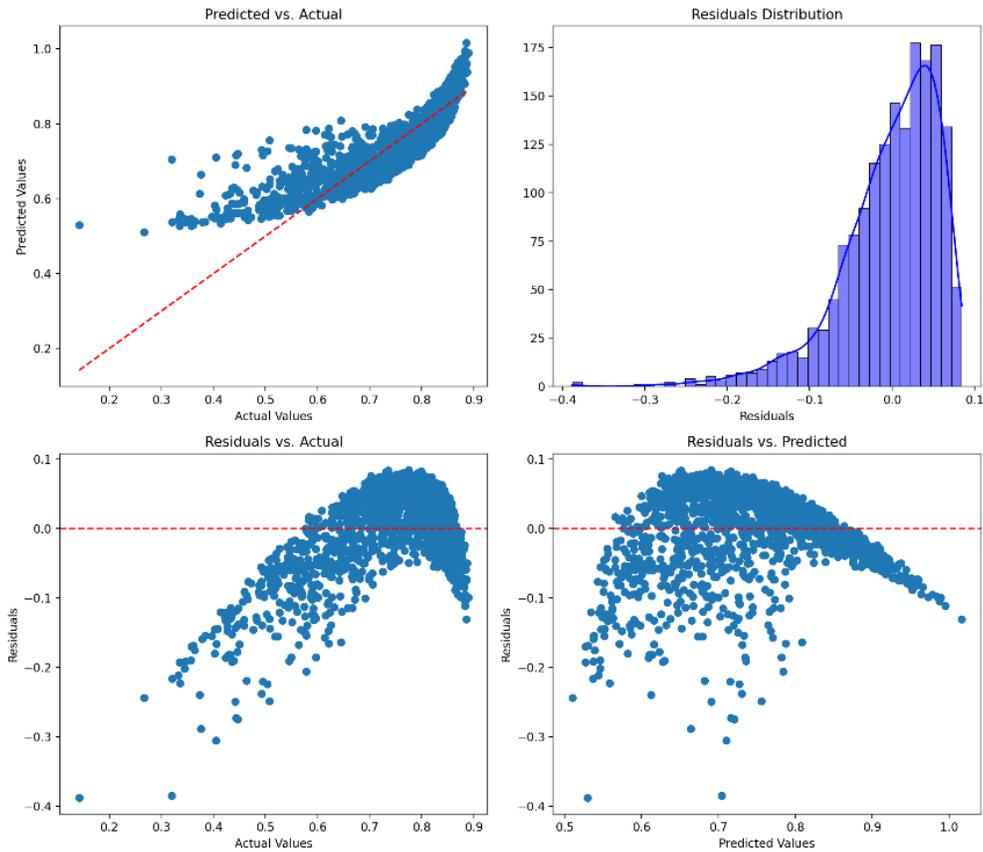


Figure 27 - Linear Regression performance, Hydrocarbon-based fuel

The simplicity of this algorithm is reflected in its poor performance evaluated through the R^2 and MSE coefficients. But what appears more clearly is its residual distribution compared with the Actual and the Predicted values which, in both cases, shows a parabola-shaped pattern which is a symptom of a bad regression. That can be addressed to the higher complexity of the SOFC model which cannot be approximated by a linear function.

7.1.2 BAYESIAN RIDGE REGRESSION

The Bayesian Ridge Regression, performed with the command `linear_model.BayesianRidge()`, offers a probabilistic prediction of the continuous output. In this model the output is calculated via a linear combination of the inputs, the latter are treated considering their probabilistic distribution which can account for prior knowledge of the input data. This allows to calculate the probabilistic distribution of the output as well as the model uncertainty at the cost of a slight increase in the computing time [11] [12] [13].

As it is possible to assess from the performances shown in *Figures 28* and *29*, poor regression performances were obtained due to the simplicity of the algorithm used. Also in this case, a non-random pattern is present for the residuals.

Model Evaluation Plots for Bayesian Regression
Cell Voltage - Humidified Hydrogen
 $R^2 = 0.7184$ MSE = 0.001623

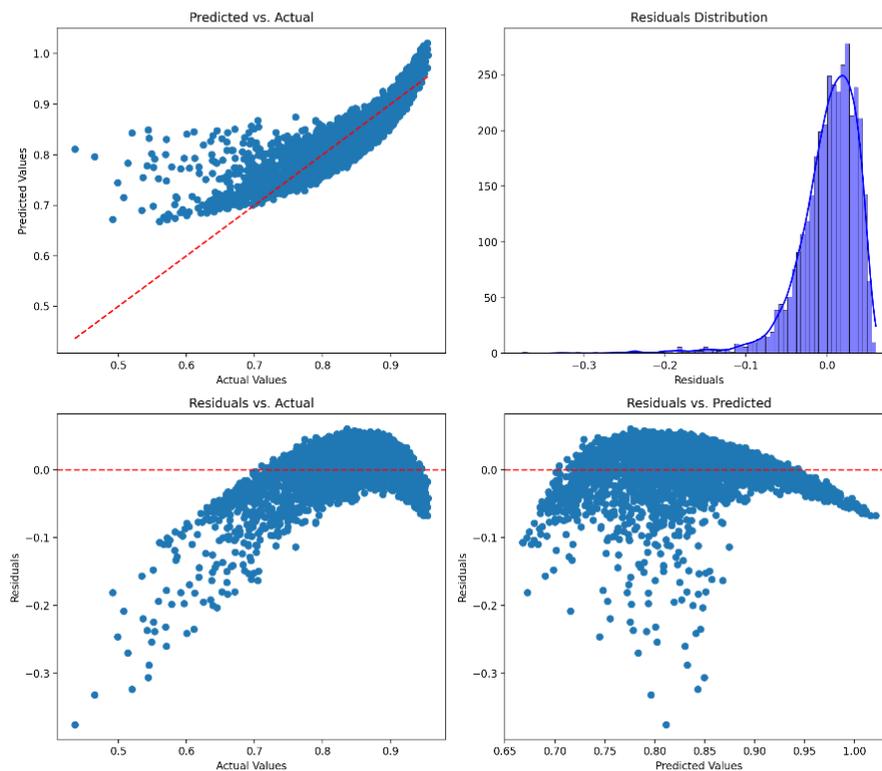


Figure 28 - Bayesian Ridge Regression performance, Humidified Hydrogen

Model Evaluation Plots for Bayesian Regression
 Cell Voltage - Hydrocarbon-based fuel
 $R^2 = 0.6975$ $MSE = 0.003795$

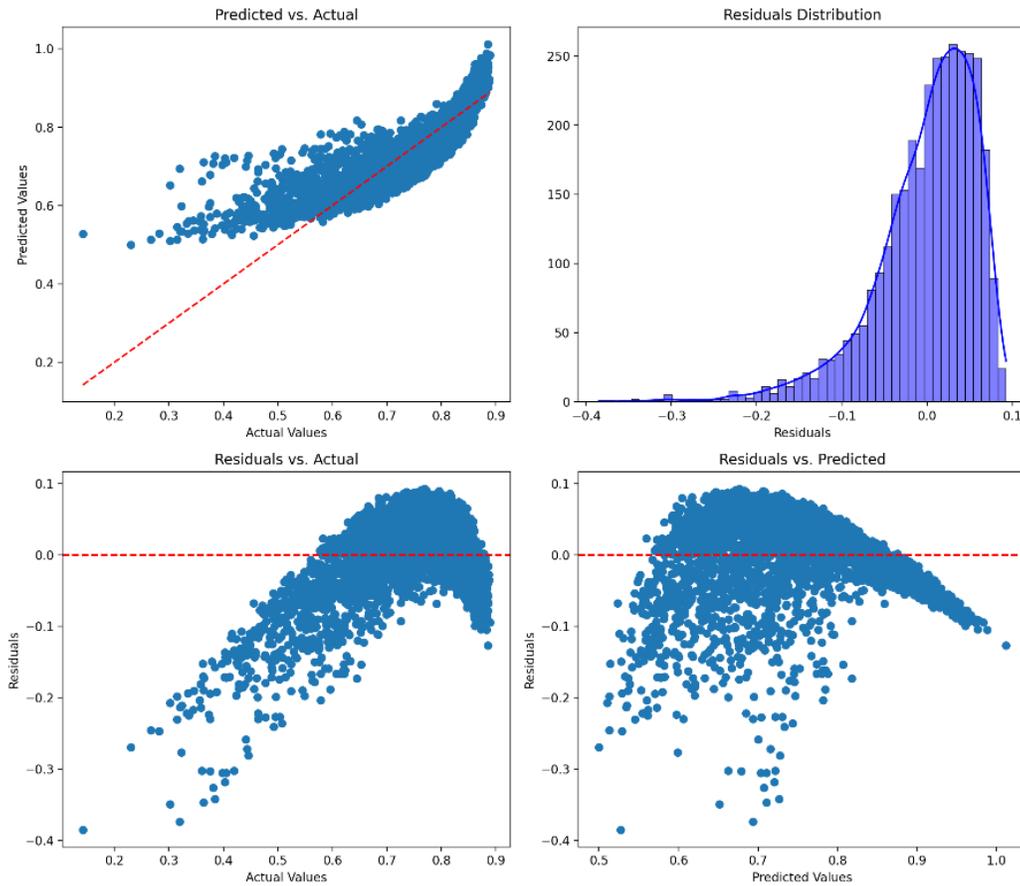


Figure 29 - Bayesian Ridge Regression performance, Hydrocarbon-based fuel

7.1.3 POLYNOMIAL REGRESSION

Polynomial Regression algorithm interpolates the data through a higher-order polynomial function which includes all the higher-interaction terms. An example of a third-degree interpolating function can be shown in the equation below, which is defined for two inputs x_1, x_2 :

$$y = f(x_1, x_2) = m_1x_1^3 + m_2x_2^3 + m_3x_1^2x_2 + m_4x_1x_2^2 + \dots + k$$

Where:

- y is the continuous output function.
- x_1, x_2 are the discrete inputs with their relative degree.
- $x_1^2 x_2, x_1 x_2^2$ are the interaction terms of the input variables.
- m_1, \dots, m_{2d+1} are the coefficients of the polynomial function.
- k is the intercept of the function.

The number of terms and coefficients of the function is directly proportional to its number of inputs. Therefore, during the implementation in Python, the function *PolynomialFeatures(d=2,3)* is used to evaluate all the interaction terms as well as the input terms with a degree higher than one. Finally, the best coefficients and the intercept, that minimize the residuals, can be calculated by recalling the command *LinearRegression()* [11] [12] [13].

It is important to underline that in theory every degree can be chosen for the polynomial function shape, but degrees higher than 4 can induce Overfitting of the data. Hence, the analysis was conducted with a polynomial degree equal to 2 and 3 and the respective results are shown in the next *Figures 30* and *31*.

Model Evaluation Plots for Polynomial Regression Degree = 2
Cell Voltage - Humidified Hydrogen
 $R^2 = 0.8902$ $MSE = 0.000642$

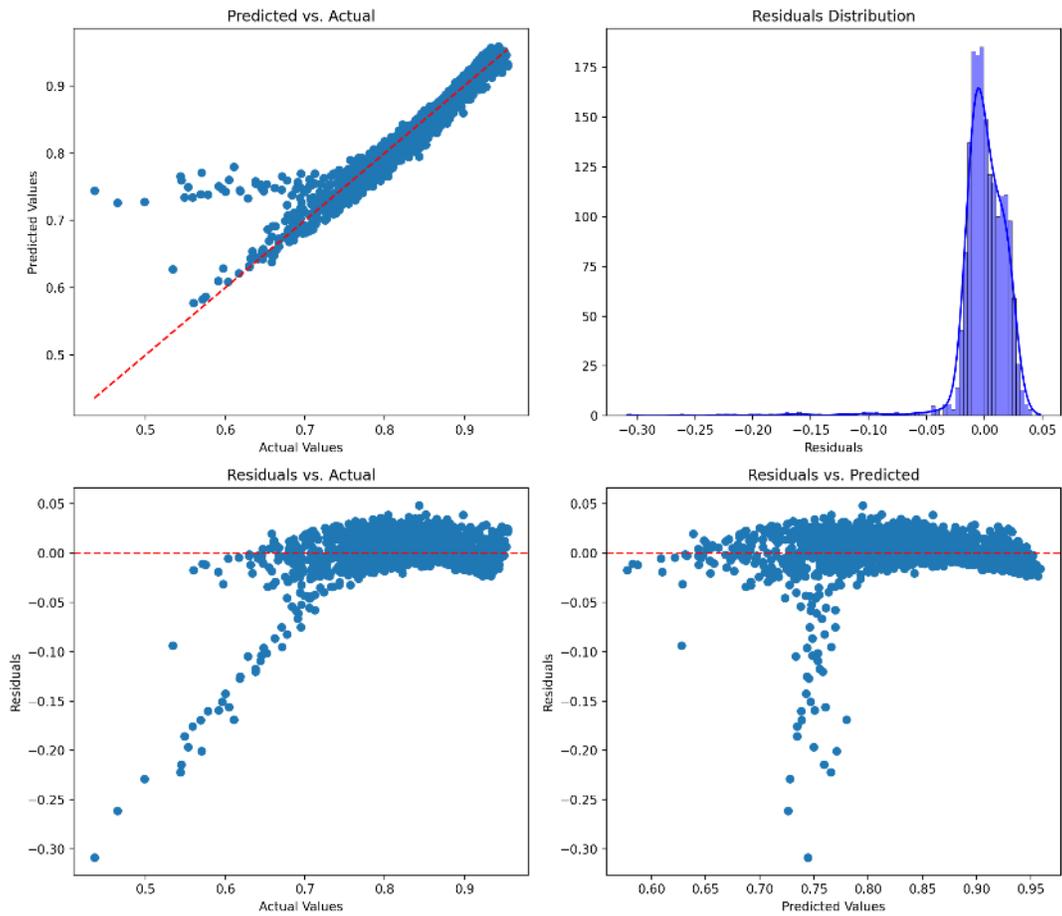


Figure 30 - Polynomial Regression $d=2$, Humidified Hydrogen

Model Evaluation Plots for Polynomial Regression Degree = 2
Cell Voltage - Hydrocarbon-based fuel
 $R^2 = 0.9301$ $MSE = 0.000895$

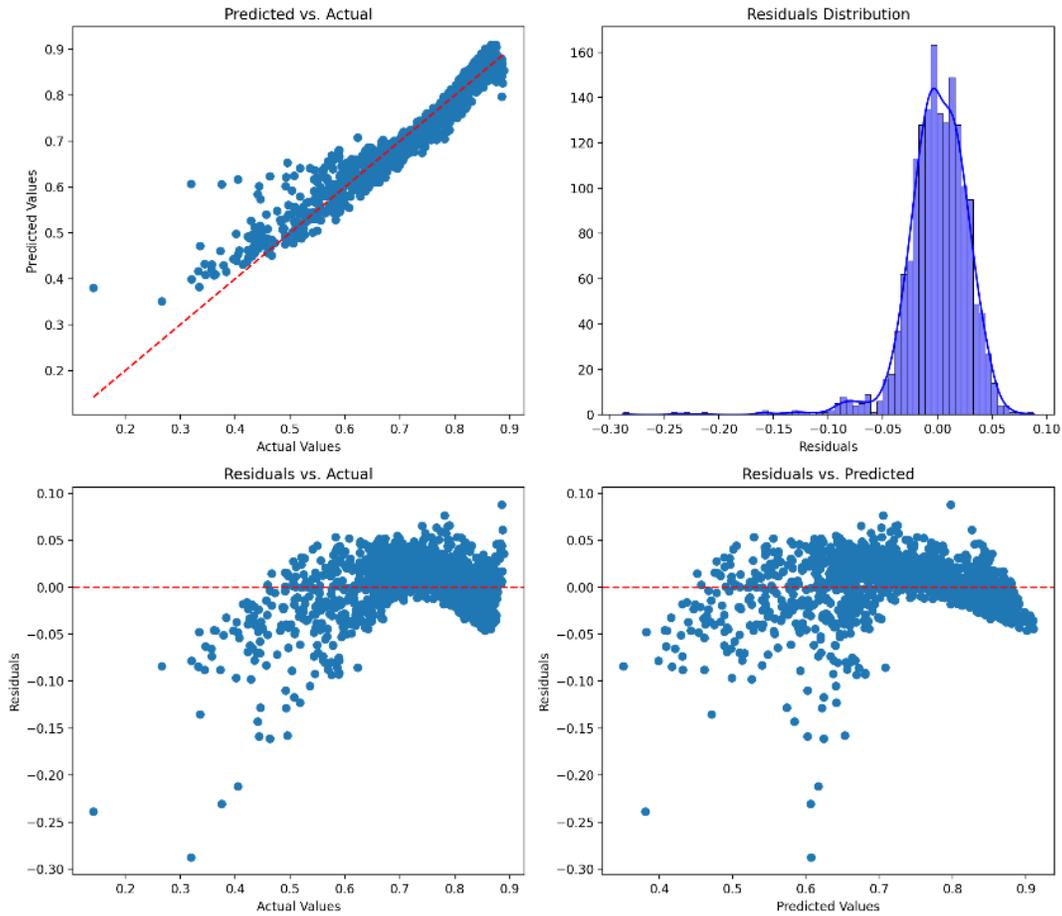


Figure 31 - Polynomial Regression $d=2$, Hydrocarbon-based fuel

The analysis of the residuals highlights the presence of non-random patterns in the residuals of both models, thus leading to extending the research to higher-order polynomial models. Figures 32 and 33 show the performance results for the Polynomial Regression with degree 3.

Model Evaluation Plots for Polynomial Regression Degree = 3
Cell Voltage - Humidified Hydrogen
 $R^2 = 0.9282$ $MSE = 0.00042$

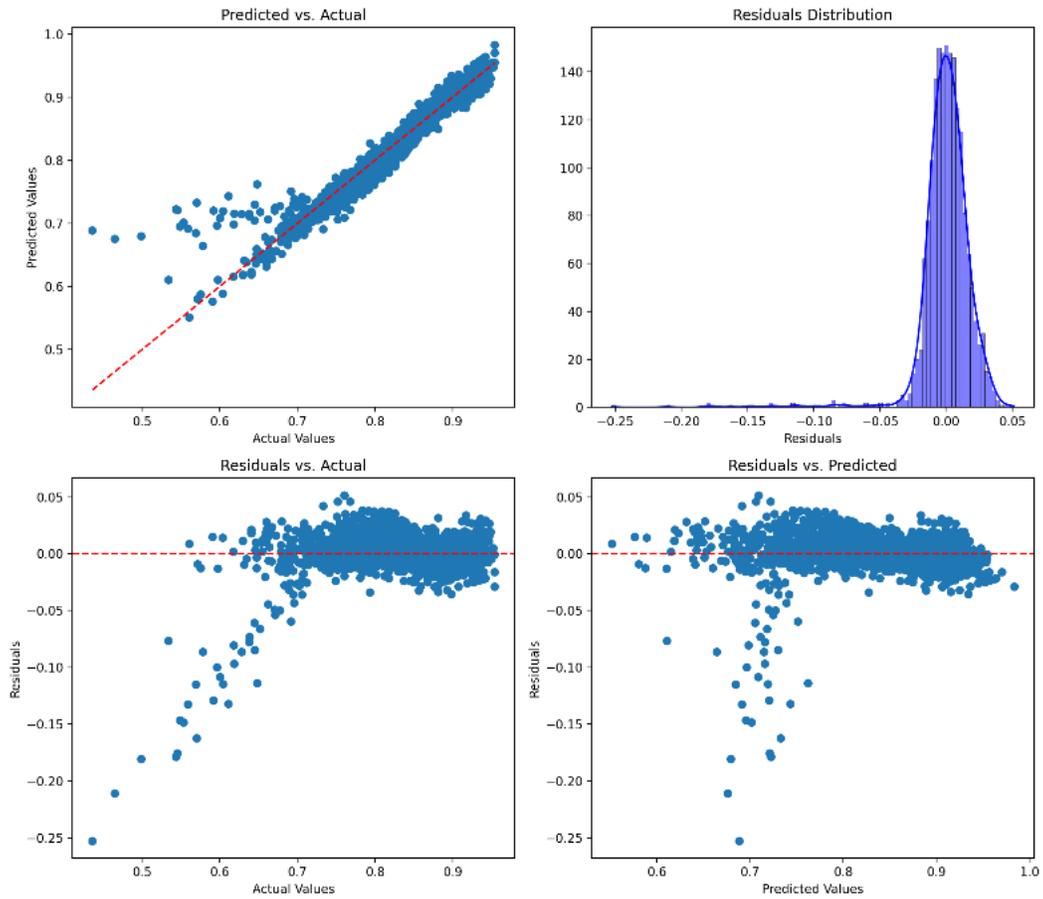


Figure 32 - Polynomial Regression $d=3$, Humidified Hydrogen

Model Evaluation Plots for Polynomial Regression Degree = 3
Cell Voltage - Hydrocarbon-based fuel
 $R^2 = 0.9699$ MSE = 0.000385

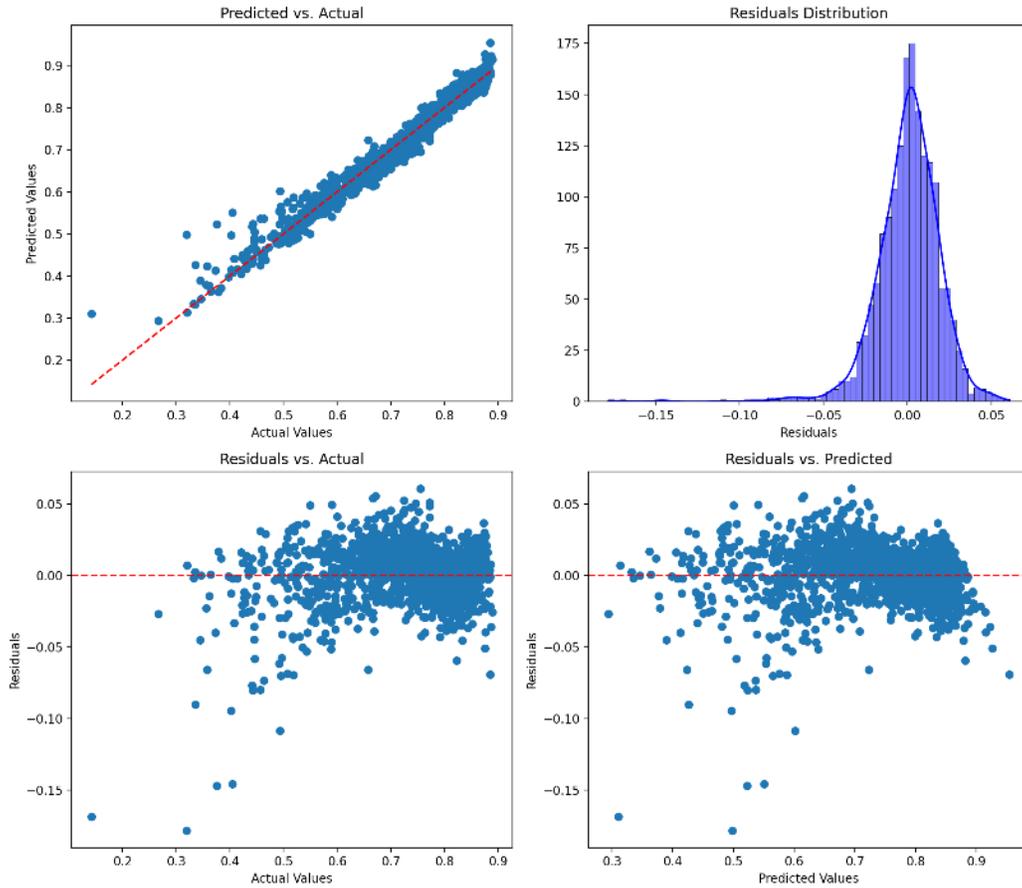


Figure 33 - Polynomial Regression $d=3$, Hydrocarbon-based fuel

The use of a third-degree polynomial led to better results for the algorithm predictions, but the analysis of the residuals still cannot be considered satisfying because:

1. In the Humidified Hydrogen the model error is quite large and more evident in the left-handed half of the analyzed voltage range.
2. Instead, in the WGSR model, the non-random pattern is more evident even if the error is widespread in the voltage range.

7.2 CLASSIFICATION ALGORITHMS

The second type of algorithms tested is the category of Classification Algorithms. These Machine Learning models have a wide range of uses, in particular, they can be employed to make predictions on discrete outputs (e.g. binary or multiple choices), make decisions or for image recognition. In this research work, a particular sub-category of the Classification Algorithms was used to predict the continuous behavior of the dependent output (i.e. Cell Voltage). The regressors tested are:

- **Decision Trees**
- **Gradient Boosting**
- **Random Forest**
- **Histogram-based Gradient Boosting**

A major difference between these algorithms and the linear models is the presence of multiple hyperparameters that can be tuned to perform better predictions. In particular, using some functions such as *GridSearchCV()* of the Scikit-Learn library is possible to automatically tune the hyperparameters to avoid problems such as Overfitting while guaranteeing the best performance in terms of R^2 and MSE [11] [12].

7.2.1 DECISION TREES

Decision Trees are the most basic models able to perform classification/regression. Due to their extreme simplicity, further algorithms were developed using the Decision Trees principle with the scope of capturing the behavior of more complex systems. *DecisionTreeRegressor* was tuned through *Grid Search Cross Validation* optimizing hyperparameters that control the size of the trees, such as `max_depth` (i.e. the maximum depth of the tree) and `max_leaf` (i.e. maximum number of nodes in the tree). In addition, the *train_test_split()* function was used to split the dataset in two: 80% dedicated to the training procedure and 20% dedicated to the testing procedure, both chosen randomly in the dataset [11] [12] [13]. The results, for both fuel models, are presented in *Figures 34* and *35*:

Model Evaluation Plots for Decision Tree Regressor
Cell Voltage - Humidified Hydrogen
 $R^2 = 0.907$ MSE = 0.000543

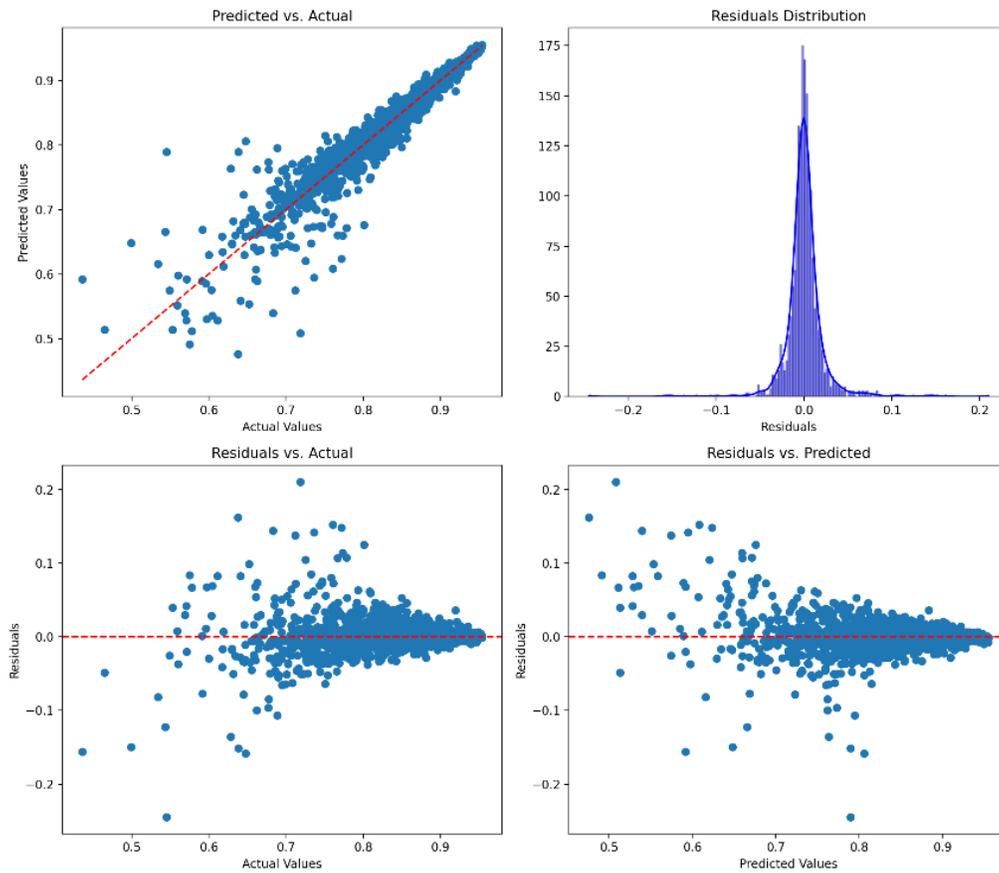


Figure 34 - Decision Trees performance, Humidified Hydrogen

Model Evaluation Plots for Decision Tree Regressor
Cell Voltage - Hydrocarbon-based fuel
 $R^2 = 0.8765$ $MSE = 0.00158$

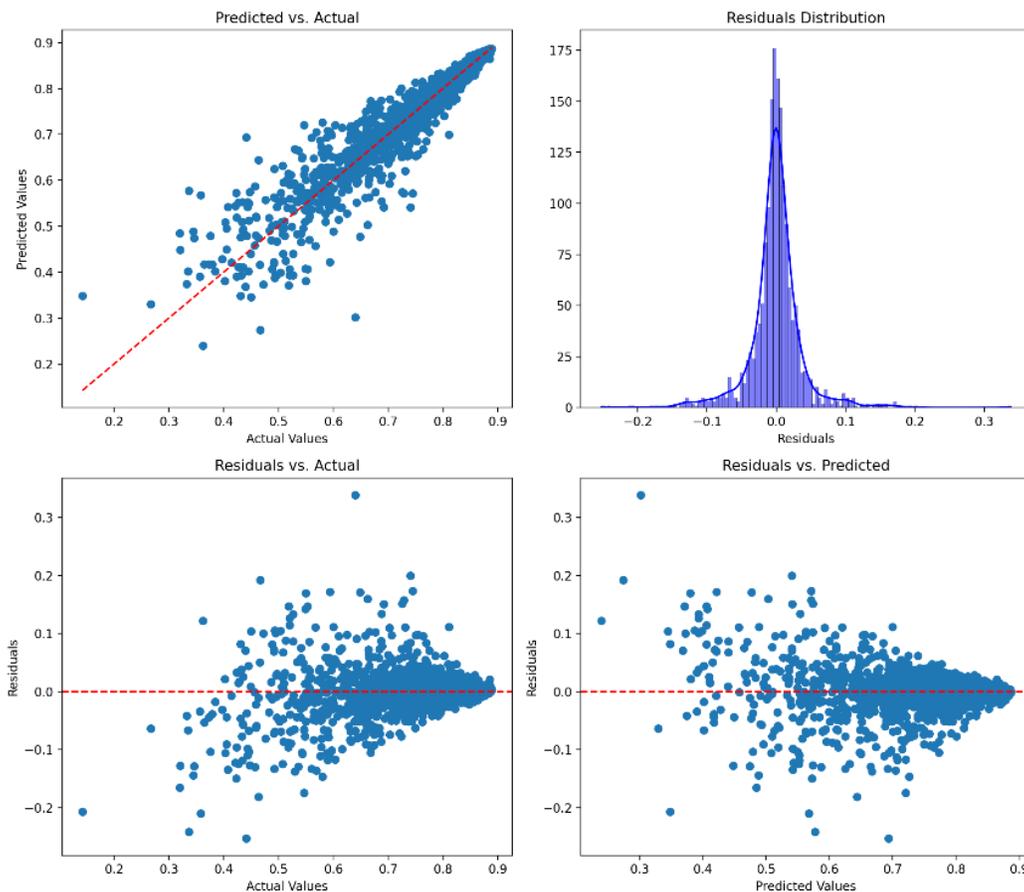


Figure 35 - Decision Trees performance, Hydrocarbon-based fuel

From the performance analysis is possible to assess that the algorithms are performing better predictions due to their residual distribution that appears to be more random and better centered concerning the linear models.

7.2.2 GRADIENT BOOSTING

GradientBoostingRegressor sets different Decision Trees in series performing the minimization of the loss function through its gradient. Different hyperparameters can

be tuned in addition to the ones related to the tree shaping: some examples are `n_estimators` (which defines the number of boosting stages to perform), `loss` (which defines the loss function type) and `learning_rate` (which shrinks the contribution of each tree avoiding Overfitting issues). Also in this case, hyperparameters were tuned with the `GridSearchCV` function and a train/test split was performed on the dataset with the specific function [11] [12]. The algorithm's performances are shown in *Figure 36* and *37*:

Model Evaluation Plots for Gradient Boosting
Cell Voltage - Humidified Hydrogen
 $R^2 = 0.9651$ $MSE = 0.000204$

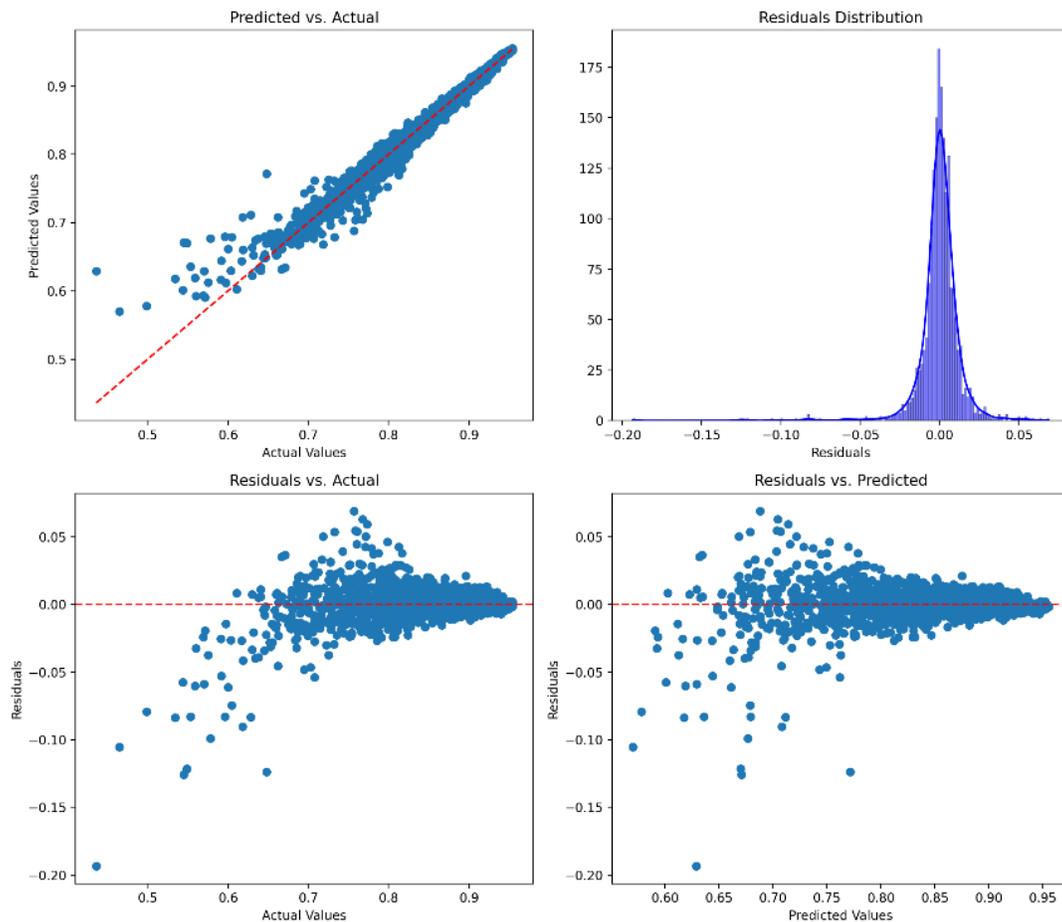


Figure 36 - Gradient Boosting performance, Humidified Hydrogen

Model Evaluation Plots for Gradient Boosting
Cell Voltage - Hydrocarbon-based fuel
 $R^2 = 0.9447$ $MSE = 0.000708$

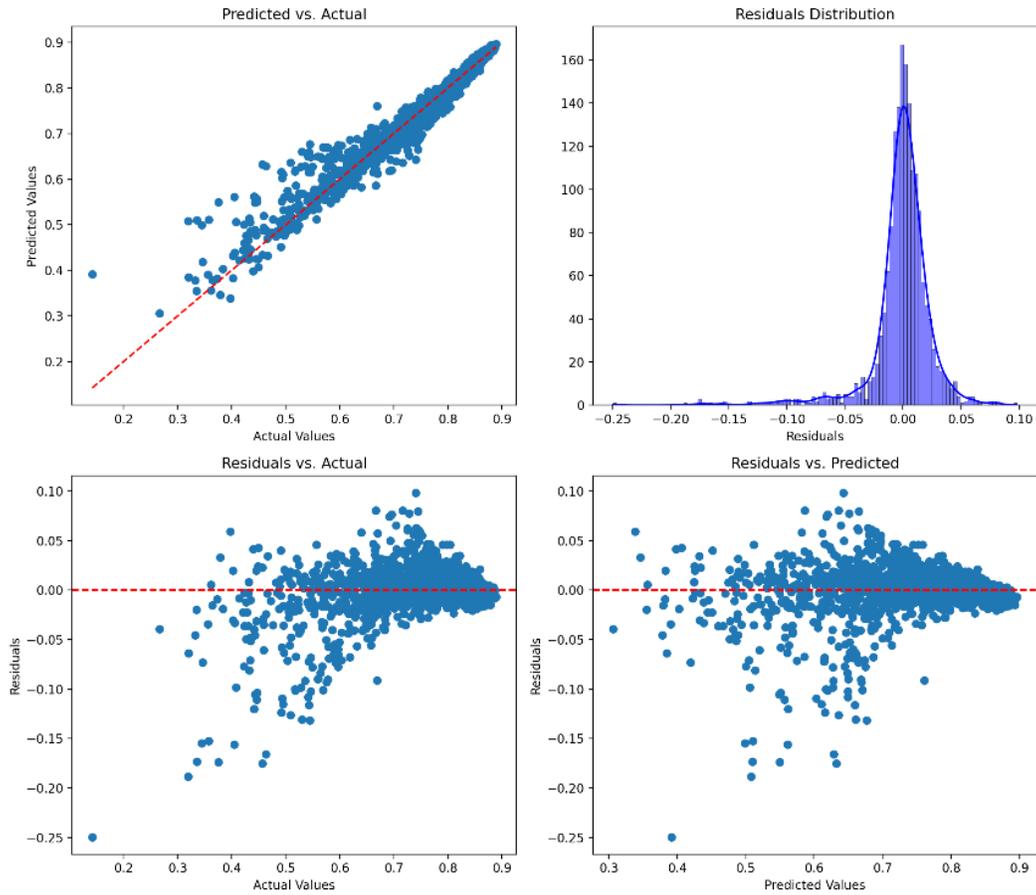


Figure 37 - Gradient Boosting performance, Hydrocarbon-based fuel

The results show good performances in terms of accuracy and averaged squared error, although some patterns can be recognized in the region included in between the range 0,15/0,65 V that results to be too large to have good applicability of the model. The reason behind this could be the large amount of data fed to the algorithm. Therefore, as suggested by the user's guide of Scikit-Learn, *Histogram-based Gradient Boosting* could be a solution and for this reason, it will be tested in the following subchapters.

7.2.3 RANDOM FOREST

RandomForestRegressor uses a multiplicity of individual Decision Trees in a parallel configuration, inhibiting the mutual interference between the trees during “the forest” generation. The algorithm is considered a meta-estimator since it averages all the decisions taken by all the individual trees. The hyperparameters are set to tune the forest generation, while Overfitting is prevented by the fact that each tree takes a random sample of the train set to build its structure (e.g. branches, leaves), thus reducing correlations within the estimators in the forest. The hyperparameters are tuned with *RandomizedsearchCV()* which finds the best values for *max_depth* and *n_estimators* that shape the forest by determining respectively the depth of each tree and the relative number of trees that ran in parallel [11] [12] [14]. The results for the Random Forest algorithm are presented below, in *Figures 38* and *39*:

Model Evaluation Plots for Random Forest Regressor
Cell voltage - Humidified Hydrogen
 $R^2 = 0.9567$ $MSE = 0.000253$

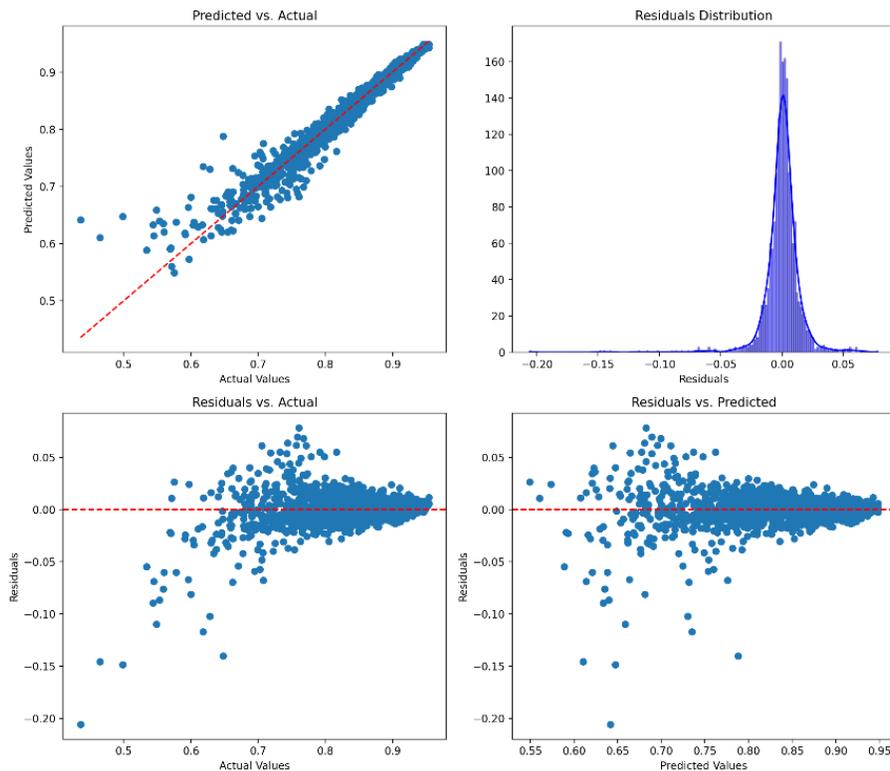


Figure 38 - Random Forest performance, Humidified Hydrogen

Model Evaluation Plots for Random Forest Regressor
Cell voltage - Hydrocarbon-based fuel
 $R^2 = 0.9346$ $MSE = 0.000837$

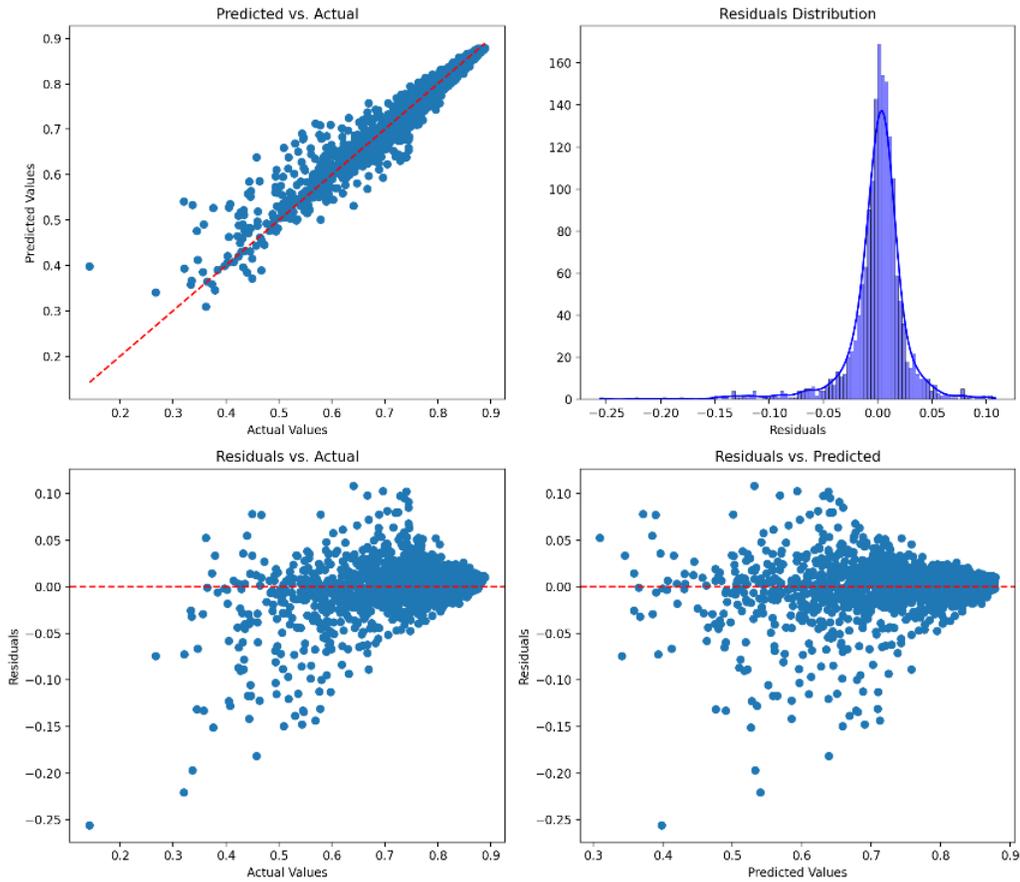


Figure 39 - Random Forest performance, Hydrocarbon-based fuel

Results show better predictions for the Humidified Hydrogen model, even if the residual distribution shows some pattern-based deviation in the left part of the evaluation range (i.e. from 0,2 to 0,7 V). In the WGSR model same pattern is present, but the cloud is more widespread from the zero-error line.

7.2.4 HISTOGRAM-BASED GRADIENT BOOSTING

This algorithm is optimized for medium-large databases (e.g. > 9500 lines) because it uses “binning” (i.e. the principle of the histogram plots) to reduce the number of trees sequentially generated during the training procedure. The hyperparameters can be tuned by *GridSearchCV* and mostly define the shape of the trees. *Learning_rate* hyperparameter coupled with *StandardScaler()* were used to prevent Overfitting [11] [12] [14]. The results are shown in the following *Figures 40* and *41*:

Model Evaluation Plots for HistGradient Boosting Regressor
Cell Voltage - Humidified Hydrogen
 $R^2 = 0.9728$ $MSE=0.000159$

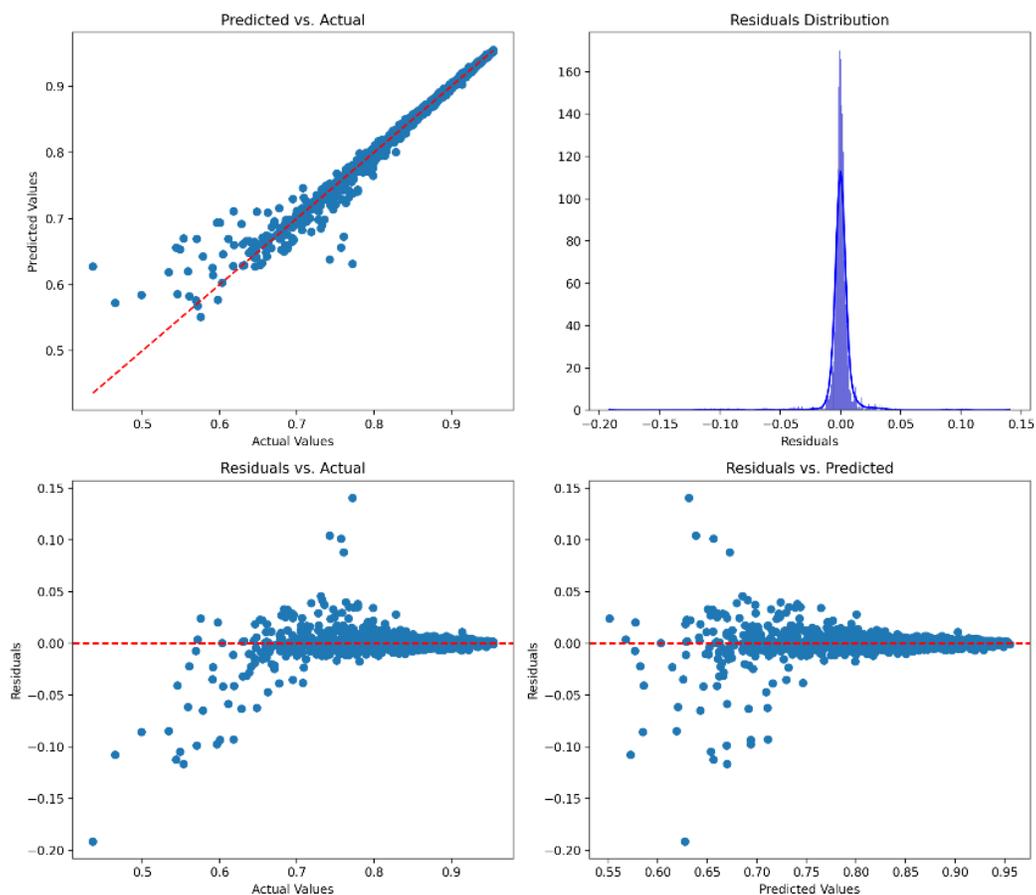


Figure 40 – HistGradientBoosting performance, Humidified Hydrogen

Model Evaluation Plots for HistGradient Boosting Regressor
Cell Voltage - Hydrocarbon-based fuel
 $R^2 = 0.9716$ $MSE=0.000364$

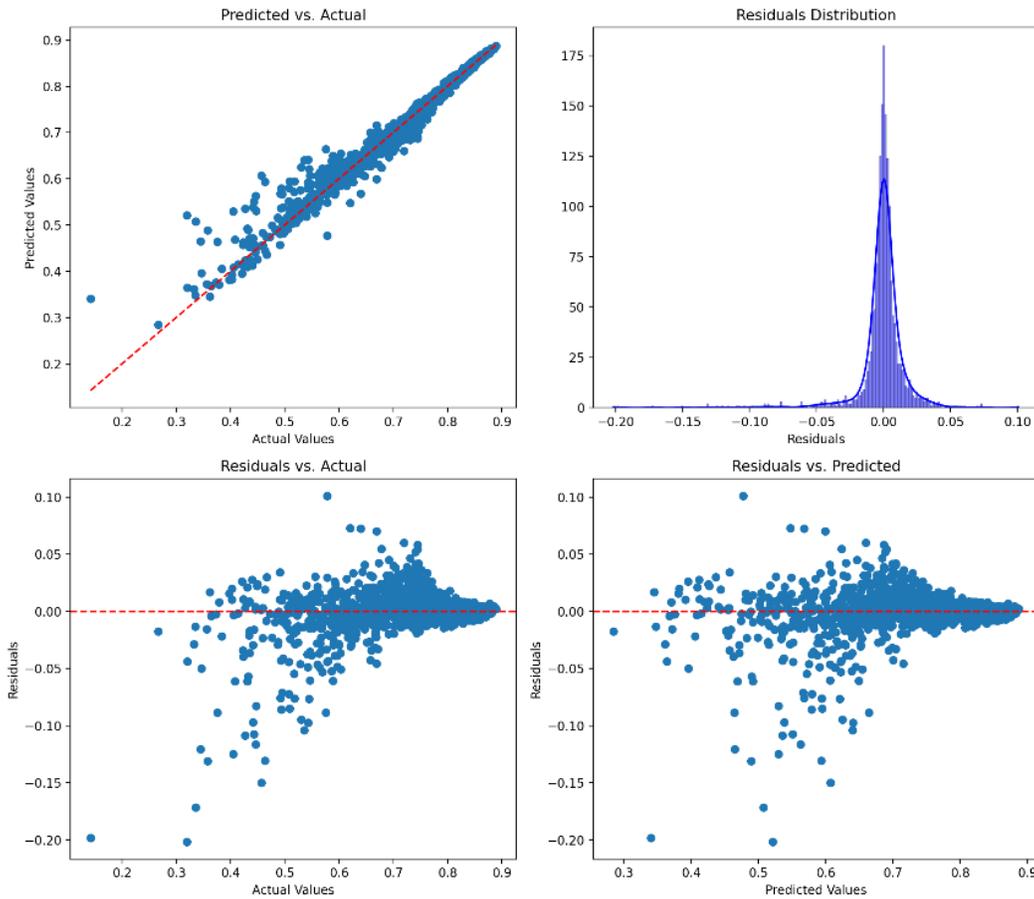


Figure 41 - HistGradientBoosting performance, Hydrocarbon-based fuel

So far, this algorithm showed the best values for the performance indicator R^2 and the lowest average squared error MSE for both the fuel models. Nevertheless, from the residual analysis is possible to notice the presence of the same pattern on the left-hand side of the Voltage range explored, thus requiring ulterior data analyses for better understanding.

7.3 ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (*ANNs*) are known to be the most flexible and the fastest high-level algorithm used for Machine Learning and Deep Learning. ANNs use neurons organized in layers, also called “nodes”, which can be represented by a regression model as a standalone. Each node is composed of input data, weights, output data, and a threshold. Once the input is defined, the weights are assigned, and the node calculates the output. The output of the node is then processed by the activation function, if its value exceeds the threshold fixed by the “bias”, the information is then passed to the subsequent node which uses the data as an input. Concerning the previous regression algorithms, the main difference is the possibility of determining multiple outputs at the same time by imposing the shape of the output layer during the network tuning step. This is the basic working principle of a feedforward network (or *Multi-Layer Perceptron*) that can be recalled in *Python* by the command *MLPRegressor()* of the *Scikit-Learn* library. The latter can be tuned by multiple hyperparameters such as *regressor__hidden_layer_sizes*, which defines the number of hidden layers and neurons in each of them; *regressor__activation*, which sets the activation function for each layer; *regressor__learning_rate_init* and *regressor__max_iter*, which are imposed to avoid Overfitting. The network was tuned using *Pipeline()* which divided the dataset into different portions using a train/test split in the selected data; the data in each portion was then standardized using *StandardScaler()* and finally the best hyperparameters were searched using *GridSearchCV()* [11] [12] [13]. To have a comparison with other Machine Learning models, the performance was tested for the prediction of the Cell Voltage as a single output. In addition, the algorithm was tested for the prediction of all four outputs, i.e. Cell Voltage, Ohmic, Activation and Concentration losses.

Model Evaluation Plots for Neural Network
Cell Voltage - Humidified Hydrogen
 $R^2 = 0.8094$ $MSE = 0.001128$

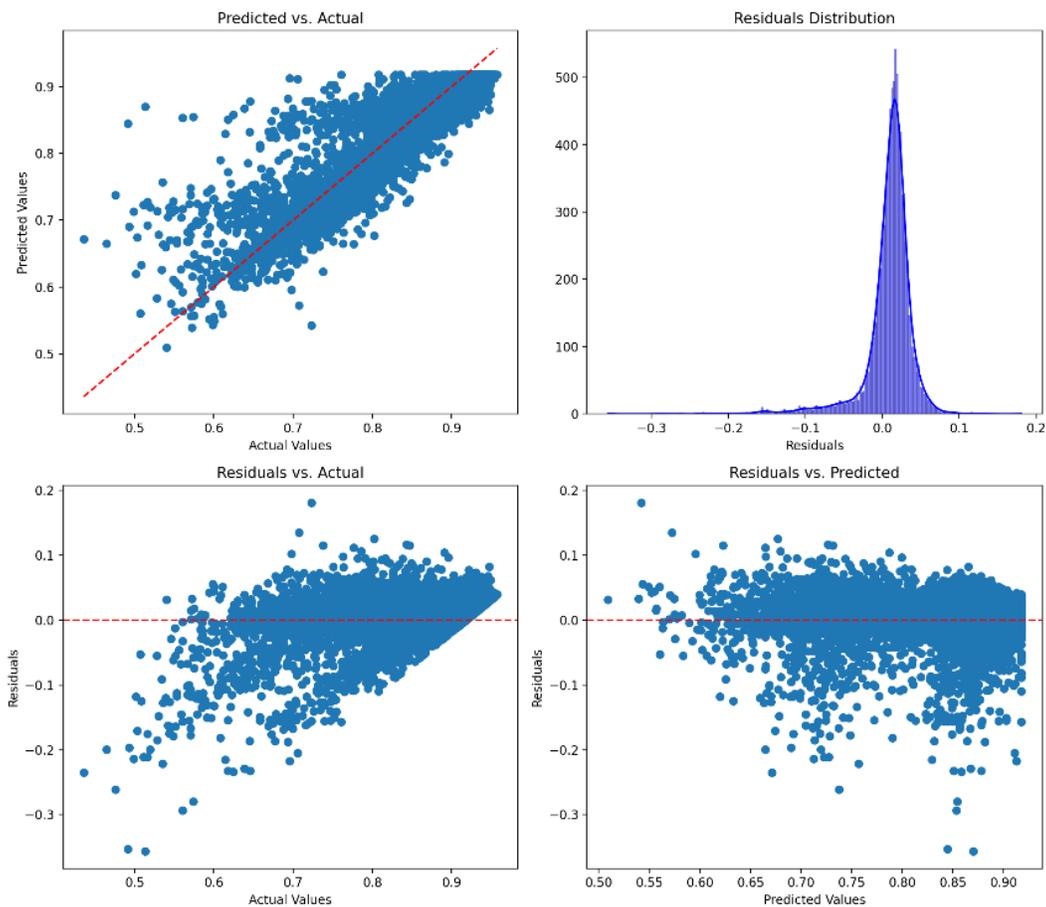


Figure 42 - ANN performance, Cell Voltage, Humidified Hydrogen

Figure 42 presents the results obtained from a two-hidden layers network made by 12 and 3 neurons in each layer. The algorithm works with 18 input neurons and 1 output neuron for the Cell Voltage prediction. The residual analysis shows a pattern-based distribution and confirms the poor prediction performances evaluated with R^2 parameter.

Model Evaluation Plots for Neural Network
 Cell Voltage - Hydrocarbon-based fuel
 $R^2 = 0.7998$ MSE = 0.002398

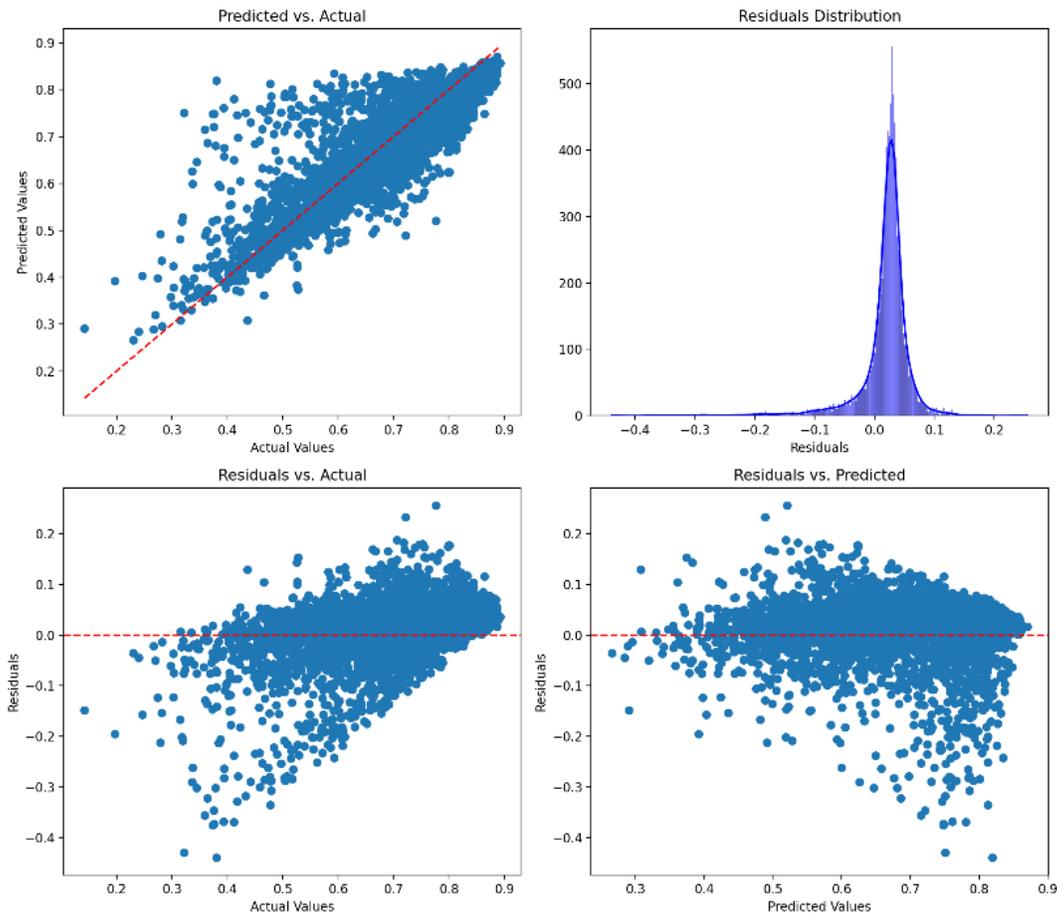


Figure 43 - ANN performance, Cell Voltage, Hydrocarbon-based fuel

Figure 43 shows the results obtained for the WGSR model, with a three hidden layers network made by respectively 16, 10 and 4 neurons in each layer. Also, this neural network works with 18 input neurons and 1 output neuron for the Cell Voltage prediction. The residual analysis shows the same pattern-based distribution as the Humidified Hydrogen model. In the following figures are presented the results for the 4-outputs ANNs.

Model Evaluation Plots for Neural Network
All outputs - Humidified Hydrogen
 $R^2 = 0.8445$ $MSE = 0.000319$

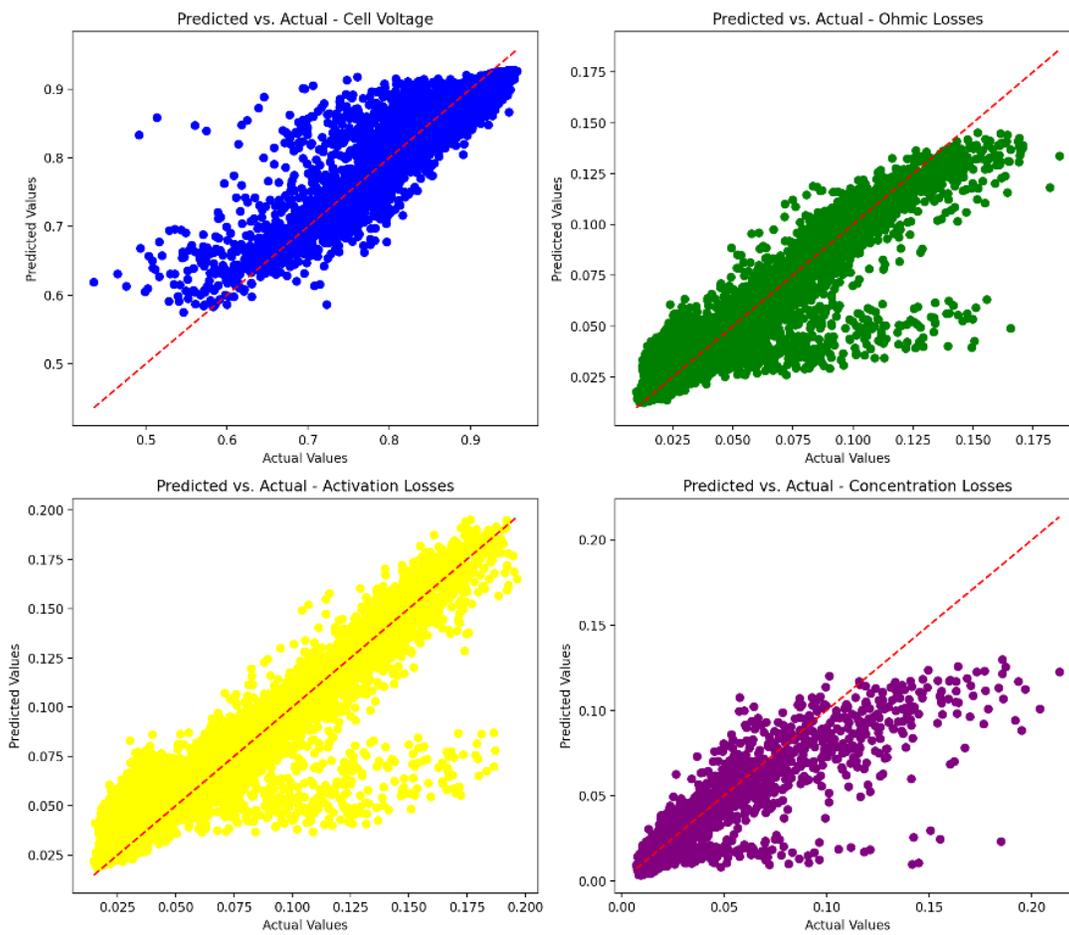


Figure 44 - ANN performance, Humidified Hydrogen, Voltage and Losses

Model Evaluation Plots for Neural Network
 All outputs - Hydrocarbon-based fuel
 $R^2 = 0.8315$ MSE = 0.00083

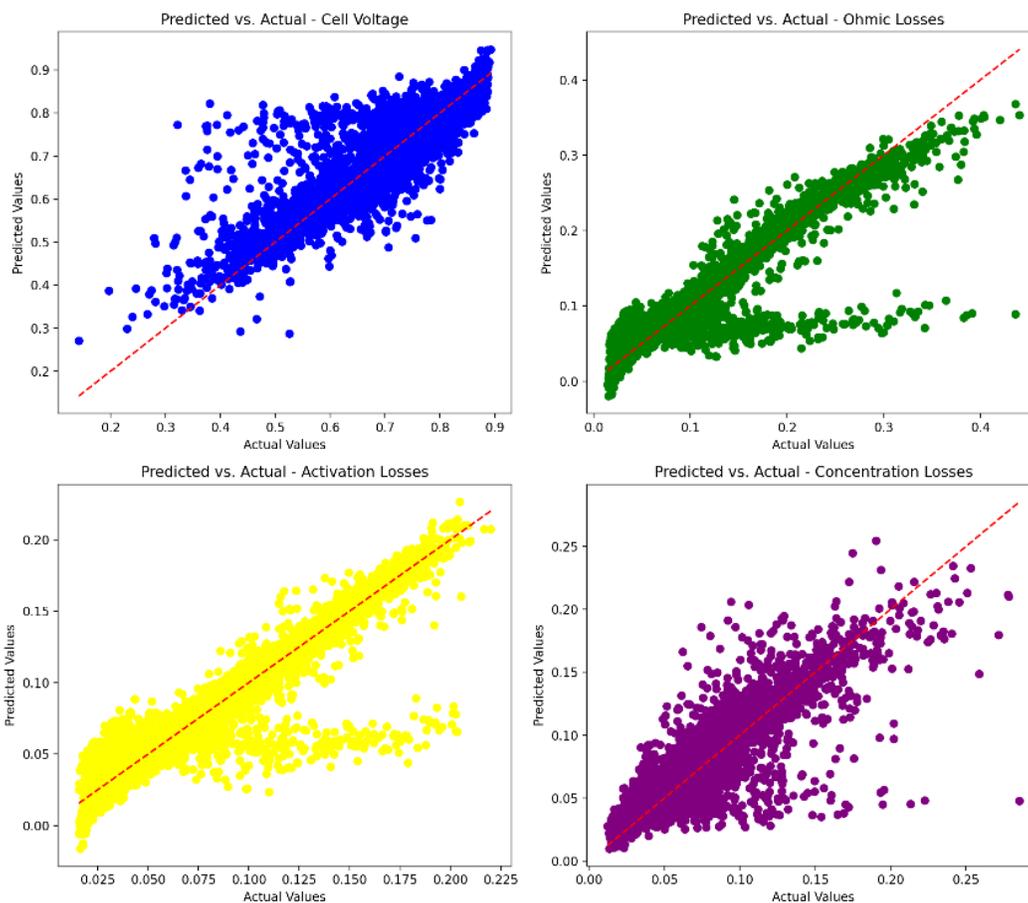


Figure 45 - ANN performance, Hydrocarbon-based fuel, Voltage and Losses

The results in the previous *Figures 44* and *45* show, surprisingly, better prediction performance with respect to the single-output case. In the first case, the network was built with three hidden layers with 24,7 and 5 neurons; instead, for the WGSR model a single hidden layer network with 36 neurons was used. Although the results are better than the single output case, the neural networks studied cannot be considered enough accurate to be employed for predictions. The reason behind these poor performances is due to the presence of more than 13 different hyperparameters that must be investigated and tested for multiple values, while during this thesis work, only 4 of them were

treated. Just regressor_hidden_layer_sizes required a trial-and-error procedure, which involved 30 different configurations to reach every result presented before. Therefore, it is obvious that, with a finer tuning of the hyperparameters, this algorithm may produce excellent results but, due to the limited time available for the experimentation, any further trial was considered unnecessary.

7.4 RESULTS

The results related to the previous analysis are listed in *Table 6*:

SOFC MODEL		Humidified Hydrogen		WGSR	
ML MODEL	Algorithm	R^2	$RMSE$	R^2	$RMSE$
LINEAR MODELS	Linear Regression	0.702	0.01742	0.707	0.00374
	Bayesian Ridge Regressor	0.718	0.00162	0.697	0.00379
	Polinomial Regression degree 2	0.890	0.00064	0.930	0.00089
	Polinomial Regression degree 3	0.928	0.00042	0.969	0.00038
CLASSIFICATION ALGORITHMS	Decision Tree	0.907	0.00054	0.876	0.00158
	Gradient Boosting Regressor	0.965	0.00020	0.944	0.00070
	Random Forest Regressor	0.956	0.00025	0.934	0.00083
	Histogram-based Gradient Boosting	0.972	0.00016	0.971	0.00036
NEURAL NETWORKS	Multi-Layer Perceptron (single output)	0.809	0.00112	0.800	0.00239
	Multi-Layer Perceptron (4 outputs)	0.844	0.00032	0.831	0.00083

Table 6 - ML algorithms performance results

From the results presented in *Table 6*, it is possible to assess that Classification Algorithms returned the best prediction accuracies and the lowest errors. Anyway, the analysis of the residuals showed that the best option among the algorithms is *Histogram-based Gradient Boosting* since the error is lower and randomly spread across all the Voltage range considered.

7.5 MODEL APPLICABILITY

The outcomes of the analyses carried out in this section prove that *Histogram-based Gradient Boosting* is the best algorithm to employ for Cell Voltage behavior predictions. The following step of the modeling procedure is the evaluation of the so-called Model Applicability. This study is performed to evaluate the extent to which a specific model can be effectively used to make predictions or take decisions in particular scenarios. The procedure involves an in-depth overlook of the residuals, i.e. the errors, to define safe limits that the scientist can apply to the model without incurring large results uncertainty. The most immediate approach is to compare the prediction error with the predictors, i.e. the input parameter or independent variables of the model. The results for the *HGB* algorithm are presented in *Figures 46* and *47*.

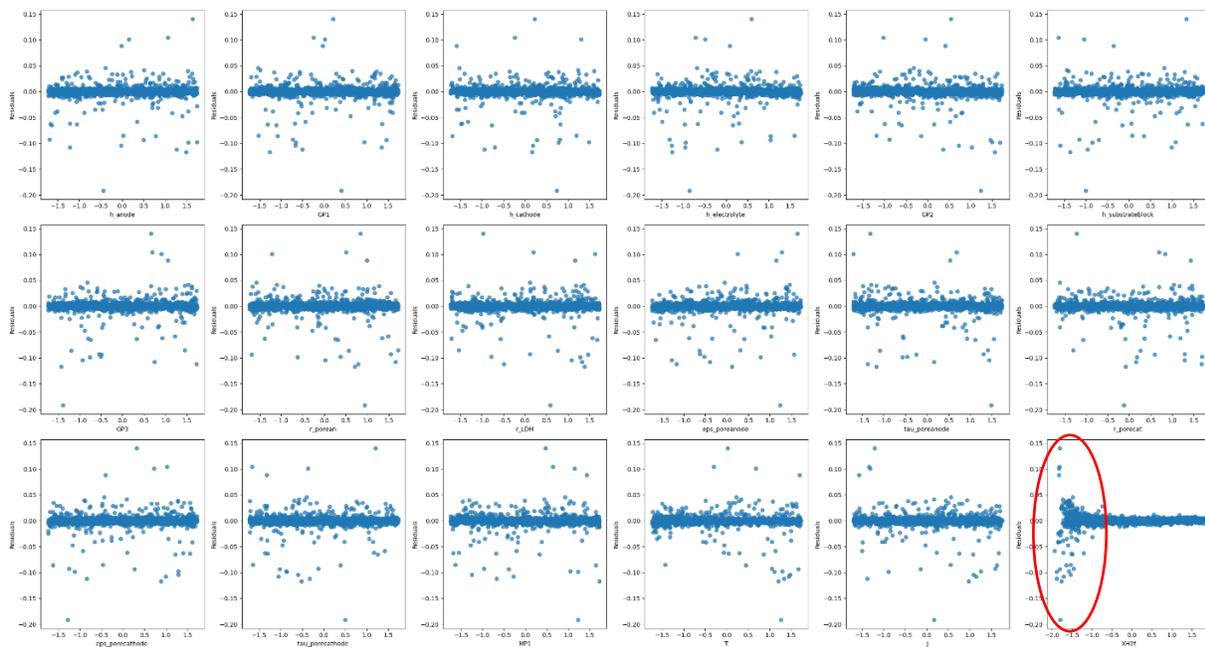


Figure 46 - HGB residual distribution, Humidified Hydrogen

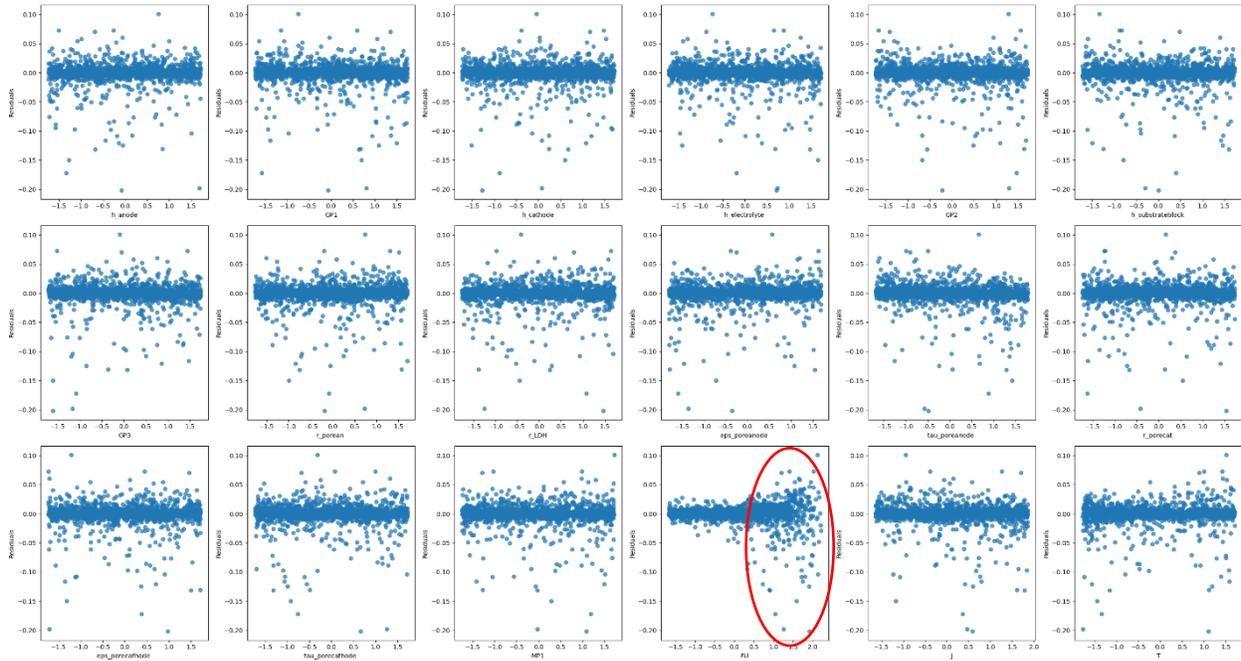


Figure 47 - HGB residuals distribution, Hydrocarbon-based fuel

Considering the plots presented above, is possible to evaluate the magnitude of the error based on the value of the independent variable chosen. It is clear, just by looking at the images, that most of the outliers are randomly spread on the various parameter ranges, except for the Hydrogen molar fraction χ_{H_2} and the Fuel Utilization factor FU . In the first case, too low concentration of Hydrogen will induce larger prediction errors. In the second case, the relation between FU and the error is directly proportional since FU is defined as:

$$FU = \frac{n_{H_2,provided}}{n_{H_2,used}}$$

Therefore, confirming that the lower the Hydrogen concentration, the larger the resulting error. These assumptions can also be confirmed by carrying out an analysis of the algorithm input dataset with the scope of investigating the causes of the “NaNs” resulting from the *COMSOL Multiphysics* simulations. *Figure 48* was produced using the *Pair-Plot* command in *Python* and highlights the statistical frequency of the parameters that led to a successful simulation (Run = 1) and vice versa (Run = 0). As it is possible to notice, the abnormal distribution is only linked to the hydrogen

concentration and the frequency is higher when the concentration is close to zero, resulting in a compiler error. Therefore, it is safe to assess that, for lower concentrations of hydrogen, the outcomes of the Multiphysics simulations were already affected by a significant linear error, thus causing poor interpolation from the ML model. Consequently, the black-box model can be safely used for Hydrogen molar fraction values close to its nominal value or higher ones, vice versa is recommended to use the Fuel Utilization Factor for values close to the nominal one or lower.

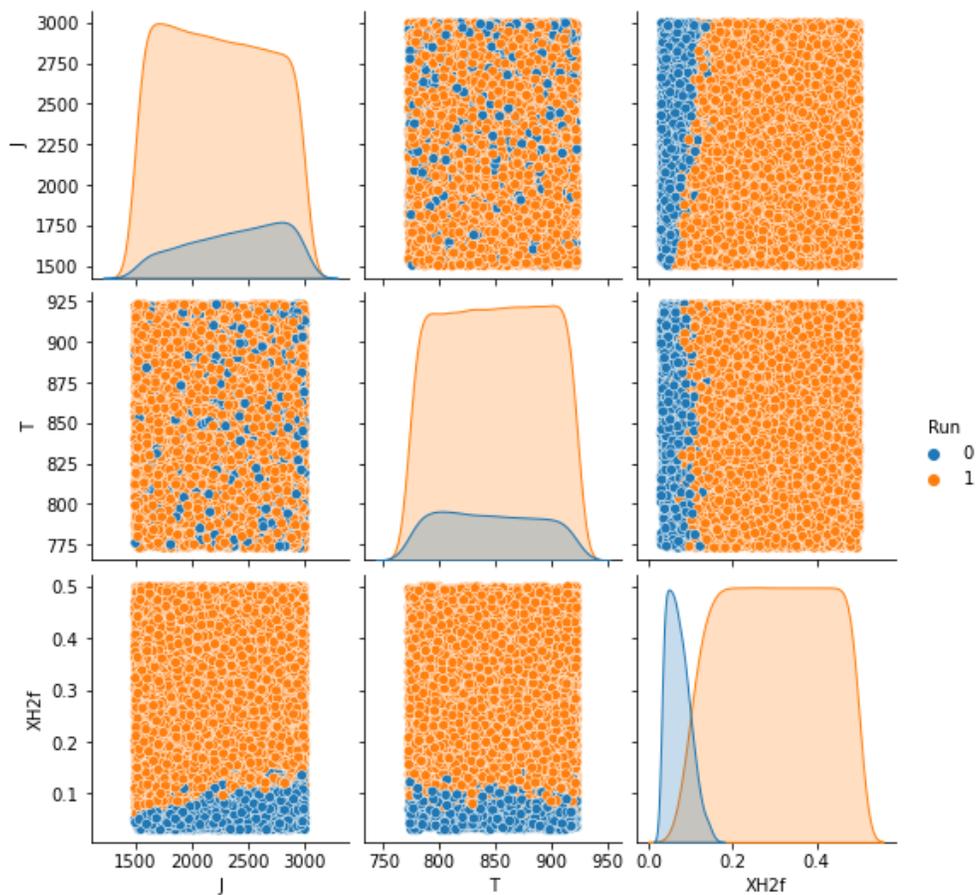


Figure 48 - Pair Plot for "NaNs" identification, Humidified Hydrogen

CONCLUSIONS

This research work was conducted on the SOFC, developed by *Robert Bosch GmbH* in a joint venture with *Ceres Power*, to retrieve a suitable Machine Learning model able to predict the cell voltage and the relative losses. To fulfill this objective, different modeling techniques were investigated to find all the important actors involved in the power generation process.

The first step involved the analysis of the physical model of the fuel cell with the identification of the process parameters, as well as the governing laws. All equations showed explicit relationships with operating conditions such as temperature, current density and fuel/air concentrations. Instead, the relationships with the design parameters were not clearly defined for all the actors involved, thus requiring further analyses. The parameter selection was based on the possibility of performing a future design improvement, while the operating conditions were investigated by selecting different points on the cell surface. After the parameter identification, a Sensitivity Analysis was performed through simulations carried out on two validated *COMSOL MultiPhysics* models, according to the type of fuel fed to the cell. The results indicated a strong influence of the anode-related parameters such as porosity, tortuosity, radius of the pores, and thickness compared to the cathode. Furthermore, the analysis revealed that the parameter sensitivity magnitude changed according to the fuel model studied. The need for a dataset to train and test the mathematical models analyzed in the final part of this thesis work led to the involvement of the Design of Experiments techniques. This gray-box modeling was used for the creation of a random dataset with 10,000 different parameter combinations that were simulated on *COMSOL* through a *Python* script. The simulation outcomes were evaluated with *ETAS-ASCMO*, the latter created a predictive model able to evaluate all the interactions between parameters as well as its own Sensitivity Analysis which confirmed previously obtained results. The dataset and the relative outcomes were used to train and test different Machine Learning algorithms such as Linear Models, Classification algorithms and Neural Networks. Linear models proved to be too simple to correctly predict the cell behavior, thus resulting in large prediction errors and bad residuals distribution. Neural Networks instead, were governed by a multiplicity of hyperparameters not fully investigated due to the lack of available time. In total 120 network shapes were tested, but the results never crossed 82% of accuracy score for the cell voltage prediction. The best option resulted in **Histogram-based Gradient Boosting**, which was able to predict the cell voltage with an accuracy score larger than 96% for both fuel models. This classification

algorithm can work with medium-large datasets and has few hyperparameters (easily tuned with a *Grid Search* function) able to increase the prediction accuracy while avoiding overfitting issues. This SOFC black-box model was further analyzed to define its application limits which resulted being Hydrogen concentration values between zero and 15%. However, the model needs to be validated through a dedicated experimental procedure.

This thesis work has been really stimulating as it has allowed to cross all the fields of modeling, from the complexity of the physical models to the new machine learning techniques. Working with complex physical systems often requires the use of mathematical models that allow you to discover otherwise invisible relationships. The latter, however, need to be read and used in a conscious way in order to best express their predictive potential, so previous knowledge is fundamental in order to understand the results obtained.

The main limits came from the usual tradeoff between performance and time that did not leave space for a more precise setting of the hyperparameters for the algorithms treated. Further research could be focused on the fine-tuning of neural networks which, if properly implemented, would lead to the creation of a single model capable of predicting multiple outputs with higher accuracy. The latter would also be able to speed up computational times, which would make it usable within loops for real-time control for a possible implementation in the fuel cell stack system. However, the produced results allowed the creation of a web application via the *Streamlit* library implemented in *Python*. This tool offers an easy and user-friendly interface to the upscaled model where the user can select the desired fuel, set any parameter value, and visualize the resultant Polarization Curves in a matter of a few seconds. Any colleague of the SOFC R&D department can access the tool from any digital device, without the need of a specific software license or prior knowledge in the field of simulation.

References

- [1] **Ember**, "Yearly Electricity Data," 2023.
- [2] **McKinsey & Co.**, 27 April 2022. [Online]. Available: <https://www.statista.com/statistics/1262346/electricity-generation-worldwide-forecast/>. [Accessed 28 October 2023].
- [3] **K. W. M. S. P. L. Jaroslaw Milewski**, *Advanced Methods of Solid Oxide Fuel Cell Modeling*, Springer, 2011.
- [4] **B. G. a. V. H. Shigenori Mitsushima**, *Fuel Cells and Hydrogen*, Elsevier, 2018.
- [5] **S.-Y. L. Khalid Zouhri**, "Exergy study on the effect of material parameters and operating conditions on the anode diffusion polarization of the SOFC," 2016.
- [6] **J. B. G. Kevin Huang**, *Solid Oxide Fuel Cell Technology*, Woodhead Publishing, 2009.
- [7] **S. T. S. Z. A. A. M.L. Faro**, "Chapter 4 - Solid oxide fuel cells," in *Compendium of Hydrogen Energy*, Elsevier, 2016.
- [8] **Q. M. i. t. B. G. |. T. S. Robert Bosch GmbH**, Booklet n.11 - Design of Experiments (DoE), Stuttgart: SOCOS, 2010.
- [9] **L. S. a. R. A. D. Solomatine**, "Data-Driven Modelling: Concepts, Approaches and Experiences," in *Water Science and Technology Library*, Springer.
- [10] **N. Matlof**, *Statistical Regression and Classification - From Linear models to Machine Learning*, CRC Press - Taylor and Francis Group, 2017.
- [11] **Pedregosa et al.**, "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [12] **Buitinck et al.**, "API design for machine learning software: experiences from the scikit-learn project," *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108-122, 2013.

- [13] **S. Raschka**, Python Machine Learning, Packt Publishing, 2015.
- [14] **M. Di Nuzzo**, Data Science e Machine Learning - Dai Dati alla Conoscenza, Kindle Ebooks, 2021.
- [15] **P. K. M. S. M. M. J. Beniak**, "Application of mathematical modelling when determining the parameters effect of biomass densification process on solid biofuels quality," MATEC, 2018. [Online]. Available: <https://doi.org/10.1051/mateconf/201816807005>.
- [16] **R. R. ., C. I. ., S. G. ., P. S. ., W. R. I. Pilatowsky**, Cogeneration Fuel Cell-Sorption Air Conditioning Systems, Springer, 2011.
- [17] **A. A. A. C. Yaneeporn Patcharavorachot**, "Electrochemical study of a planar solid oxide fuel cell: Role of support structures," vol. 177, 2007.
- [18] **M. E. A. M. & J. F. John Hennig**, "MPh-py/MPh," Zenodo, 2023. [Online]. Available: <https://doi.org/10.5281/zenodo.7749502>.
- [19] **ETAS GmbH**, ETAS ASCMO-STATIC V5.11 - User's Guide.
- [20] **J. Y. B. S. Martin Andersson**, "SOFC modeling considering electrochemical reactions at the active three boundaries," 2011.

Table of Figures

Figure 1 - Planar type (A) and Tubular type (B) fuel cells.....	4
Figure 2 - Scheme of a singular fuel cell and main reactions.....	9
Figure 3 - Types of Modeling [8]	13
Figure 4 - Linear Regression (d=1) and Polynomial Regression (d=2) comparison..	16
Figure 5 - Classification Models comparison	17
Figure 6 - Structure of an Artificial Neural Network	18
Figure 7 - Examples of DoE designs [15].....	20
Figure 8 - Polarization curve of a SOFC	23
Figure 9 - Average operating conditions across the cell length.....	27
Figure 10 - Simulation geometry	29

Figure 11 - Cathode's Domain.....	30
Figure 12 – Electrolyte's Domain.....	32
Figure 13 – Anode's Domain.....	33
Figure 14 – Substrate's Domain	34
Figure 15 - Parameter Voltage Sensitivity Humidified Hydrogen	39
Figure 16 - Parameters Voltage Sensitivity Hydrocarbon Based	39
Figure 17 - Schematic view of the Space-Filling plan for both fuel simulation models	43
Figure 18 - NDoE results Humidified Hydrogen.....	44
Figure 19 – Prediction Performance GP, Humidified Hydrogen.....	45
Figure 20 - Interaction matrix for the electrodes related parameters.....	46
Figure 21 - Sensitivity Analysis Humidified Hydrogen	47
Figure 22 - NDoE Hydrocarbon Based fuel model	48
Figure 23 - Prediction performance GP, Hydrocarbon Based	48
Figure 24 - Interaction matrix for anode's and cathode's parameters	49
Figure 25 - Sensitivity Analysis Hydrocarbon Based fuel model	49
Figure 26 - Linear Regression performance, Humidified Hydrogen.....	53
Figure 27 - Linear Regression performance, Hydrocarbon-based fuel	54
Figure 28 - Bayesian Ridge Regression performance, Humidified Hydrogen	55
Figure 29 - Bayesian Ridge Regression performance, Hydrocarbon-based fuel.....	56
Figure 30 - Polynomial Regression d=2, Humidified Hydrogen.....	58
Figure 31 - Polynomial Regression d=2, Hydrocarbon-based fuel	59
Figure 32 - Polynomial Regression d=3, Humidified Hydrogen.....	60
Figure 33 - Polynomial Regression d=3, Hydrocarbon-based fuel	61
Figure 34 - Decision Trees performance, Humidified Hydrogen.....	63
Figure 35 - Decision Trees performance, Hydrocarbon-based fuel.....	64
Figure 36 - Gradient Boosting performance, Humidified Hydrogen.....	65
Figure 37 - Gradient Boosting performance, Hydrocarbon-based fuel	66
Figure 38 - Random Forest performance, Humidified Hydrogen	67
Figure 39 - Random Forest performance, Hydrocarbon-based fuel	68
Figure 40 – HistGradientBoosting performance, Humidified Hydrogen	69
Figure 41 - HistGradientBoosting performance, Hydrocarbon-based fuel.....	70
Figure 42 - ANN performance, Cell Voltage, Humidified Hydrogen	72
Figure 43 - ANN performance, Cell Voltage, Hydrocarbon-based fuel.....	73
Figure 44 - ANN performance, Humidified Hydrogen, Voltage and Losses	74

Figure 45 - ANN performance, Hydrocarbon-based fuel, Voltage and Losses.....	75
Figure 46 - HGB residual distribution, Humidified Hydrogen.....	77
Figure 47 - HGB residuals distribution, Hydrocarbon-based fuel.....	78
Figure 48 - Pair Plot for "NaNs" identification, Humidified Hydrogen	79
Table 1 - Fuel cell types comparison (3)	6
Table 2 - Reactions and Nernst Potential of SOFC electrochemical reactions (20)..	22
Table 3 - Parameters and ranges preliminary selection	37
Table 4 - Test cases for different fuel models	37
Table 5 – Parameter variation influence on the losses.....	40
Table 6 - ML algorithms performance results	76

Ringraziamenti

Vorrei ringraziare innanzitutto la mia famiglia, Mamma Francesca e Papá Bruno, mio fratello Matteo con mia cognata Camilla e la piccola Lavinia. Grazie per essermi stati sempre accanto nei momenti belli e in quelli brutti, sostenendomi moralmente ed economicamente in questo lungo percorso. Senza di voi probabilmente non sarei arrivato dove sono ora e non sarei stato la persona che sono oggi, a voi devo tutto.

Ringrazio la mia fidanzata Ilaria che é il mio punto di riferimento, la mia bussola da sempre in questo mare mosso che é la vita, fatta di alti e bassi. Grazie a te gli alti sono ancora piú belli e i bassi meno dolorosi e preoccupanti. Grazie per sopportarmi e supportarmi quotidianamente a casa e fuori. Spero di averti sempre con me perché sei il mio faro che illumina il cammino verso il futuro.

Vorrei ringraziare la mia famiglia acquisita, i nonni Rosanna e Cici, la mia seconda mamma Concettina e il mio fratellino Davide, gli zii Luigina, Fabio, Wilma, Marco, Caterina e Clemente per avermi sempre fatto sentire a casa e super coccolato. Un ringraziamento va anche ai cugini Francesca con Giorgio e Chiara.

Ci tengo a ringraziare tutti gli amici, partendo dagli storici Alessio, Falla, Daniele, Fabio, Federico e Monte per essere da sempre al mio fianco sin dai tempi delle elementari, conto di ripagarvi al Calema Wild. I miei compagni delle superiori Federica, Francesco, Mario, Matteo e Simone per aver reso piú leggeri i nostri percorsi universitari con qualche chiacchiera davanti ad una birra. I miei amici e colleghi mecCANIci Giuseppe, Silvia e Stefano che hanno alleggerito il carico di studio Universitario tra risate e partite a biliardino; il mio amico Marco per i tanti bei momenti passati assieme, spero di continuare a condividere questa bella amicizia con tutti voi in quel di Torino. Gli Stoccardesi Luca e Magda che sono stati consiglieri e maestri di sopravvivenza d'oltralpe, mi mancherete tanto ma so di poter contare sempre su di voi. Ringrazio anche tutti i compagni di viaggio Federica, Enrico, Luca P., Luigi, Tommaso, Elisa, Filip, Thivyan, Ankit, Dalina, Malileh e Sven, spero che la vita ci faccia incontrare spesso, magari in qualche bar per condividere ancora tante esperienze e risate. Mi mancheranno anche tutti gli amici sportivi del gruppo "Stuttgart Expats - Volleyball", spero di potervi incontrare nel campo e fuori!

Vorrei ringraziare tutti i miei docenti del Politecnico di Torino per avermi insegnato le basi dell'ingegneria e aver condiviso esperienze etiche per il lavoro e la vita. In particolare vorrei ringraziare i miei relatori, Prof. Vittorio Verda e Prof.ssa Martina Capone per avermi aiutato in questo elaborato armato di pazienza, disponibilità

e gentilezza. Ringrazio anche i miei supervisors in Bosch Andreas, Katarina e Winfried per gli insegnamenti e i preziosi consigli dispensati in questi mesi.

Ho deciso di dedicare quest'ultimo trafiletto a te Andrea perché é il piú difficile da scrivere. Il dolore della tua scomparsa é ancora troppo forte. Ti volevo ringraziare perché dal primo giorno mi hai accolto a casa tua e trattato come un figlio, facendomi sentire subito parte integrante della vostra famiglia. Ti volevo ringraziare perché sei stato un vero amico, gentile e sempre disponibile per un aiuto o una semplice chiacchiera. Ti volevo ringraziare per la fiducia e la stima che hai riposto in me in questi anni. Sei e sarai la persona piú buona che io abbia incontrato in vita mia, un gioiello che solo chi ha avuto la fortuna di conoscerti puó apprezzare. So che oggi saresti stato particolarmente orgoglioso di me, ma sono sicuro che d'ora in poi continuerai a guidarmi e proteggermi da lassú. Fai buon viaggio, ti voglio tanto bene.