USE OF MACHINE LEARNING TECHNIQUES IN THE MODELING OF AN INDUSTRIAL REACTION SYSTEM

Sarah Lilian de Lima Silva 1
Marcos Sousa Leite 2
Thalita Cristine Ribeiro Lucas Fernandes 3
Sidinei Kleber da Silva 4
Antonio Carlos Brandão de Araújo 5

ABSTRACT

Purpose: The objective of the research is to address the industrial production of hydrogen and develop the modeling and simulation of hydrogen-related industrial reaction systems using Machine Learning techniques.

Theoretical Framework: The research explores the innovation and promise of Machine Learning techniques in modeling industrial reaction systems, enabling the creation of flexible and adaptive models to deal with complexities in industrial processes.

Method/Design/Approach: The method involves the application of machine learning methods, such as linear regressions and the kriging or Gaussian process method, to develop metamodels that analyze the steps of an industrial reaction involving hydrogen.

Results and Conclusion: The results of the training and analysis have achieved satisfactory outcomes, with expected values assessed through constraint parameters for each output variable.

Research Implications: The research aims to improve real-time prediction accuracy, process variable control, and early fault detection, resulting in greater sustainability and economic efficiency in the industry.

Originality/Value: The values underpinning the research include the promotion of technological innovation, operational efficiency, and environmental sustainability, as well as valuing quality, safety, social responsibility, continuous improvement, global competitiveness, and regulatory compliance in the industry.

Keywords: Hydrogen, Machine Learning, System Modeling and Production.

USO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA NA MODELAGEM DE UM SISTEMA DE REAÇÃO INDUSTRIAL

RESUMO

Objetivo: O objetivo da pesquisa é abordar a produção industrial de hidrogênio e desenvolver a modelagem e simulação de sistemas de reação industrial relacionados ao hidrogênio, utilizando técnicas de Machine Learning.

1 Universidade Federal de Campina Grande (UFCG), Campina Grande, Paraíba, Brasil. E-mail: sarablima65@gmail.com Orcid: https://orcid.org/0009-0007-9396-451X
2 Universidade Federal de Campina Grande (UFCG), Campina Grande, Paraíba, Brasil. E-mail: marcosleitte23@gmail.com Orcid: https://orcid.org/0009-0001-7196-3598
3 Universidade Federal de Campina Grande (UFCG), Campina Grande, Paraíba, Brasil. E-mail: thalita.ribeiro@eq.ufcg.edu.br Orcid: https://orcid.org/0000-0002-8831-1393
4 Universidade Federal de Campina Grande (UFCG), Campina Grande, Paraíba, Brasil. E-mail: sidinei@eq.ufcg.edu.br Orcid: https://orcid.org/0000-0003-4627-8254
5 Universidade Federal de Campina Grande (UFCG), Campina Grande, Paraíba, Brasil. E-mail: brandao@eq.ufcg.edu.br Orcid: https://orcid.org/0000-0001-6373-5571
Referencial Teórico: A pesquisa explora a inovação e promessa das técnicas de Machine Learning na modelagem de sistemas de reação industrial, permitindo a criação de modelos flexíveis e adaptativos para lidar com complexidades em processos industriais.

Método: O método envolve a aplicação de métodos de aprendizado de máquina, como regressões lineares e o método da krigagem ou processo Gaussiano, para desenvolver metamodelos que analisem as etapas de uma reação industrial envolvendo o hidrogênio.

Resultados e Conclusão: Os resultados dos treinamentos e análises alcançaram resultados satisfatórios, com valores esperados avaliados por parâmetros de restrição para cada variável de saída.

Implicações da Pesquisa: A pesquisa busca melhorar a precisão das previsões em tempo real, o controle das variáveis de processo e a detecção precoce de falhas, resultando em maior sustentabilidade e eficiência econômica na indústria.

Originalidade/Valor: Os valores que fundamentam a pesquisa incluem a promoção da inovação tecnológica, a eficiência operacional e a sustentabilidade ambiental, além de valorizar a qualidade, a segurança, a responsabilidade social, a melhoria contínua, a competitividade global e a conformidade regulatória na indústria.

Palavras-chave: Hidrogênio, Machine Learning, Modelagem de Sistemas e Produção.

1 INTRODUCTION

The technological revolution of recent decades has significantly shaped how industries operate and optimize their processes. One of the most notable transformations in this scenario is the rise of Machine Learning as a powerful tool to enhance the efficiency, accuracy, and sustainability of industrial operations. Hydrogen production plays a crucial role in the quest for cleaner and more sustainable energy sources worldwide. As one of the most abundant elements in the universe, hydrogen has the potential to become a key component in the transition to a greener and more efficient energy future. However, hydrogen production is not a simple task, requiring highly complex and effective industrial reaction systems. In this context, the use of Machine Learning techniques emerges as a revolutionary tool in the modeling and optimization of these systems. This article explores the application of Machine Learning techniques in the modeling of industrial reaction systems, highlighting the benefits and possibilities that this approach offers. Modeling industrial reaction systems is a complex and essential challenge to ensure the safety, quality, and efficiency of production processes. In this context, Machine Learning emerges as a promising solution, enabling the creation of more accurate and adaptable models capable of dealing with the inherent variability of industrial systems.

2 THEORETICAL REFERENCE

2.1 Reactions for Hydrogen Production

According to Davis (2010), around 95% of hydrogen production in the United States in the mid-2010s was carried out through the process known as steam methane reforming (SMR). In this process (equation 1), methane (CH₄) from natural gas chemically reacts with extremely hot water vapor (H₂O), typically ranging between 700°C and 1000°C, producing a gas mixture composed of hydrogen (H₂) and carbon monoxide (CO), as described by Davis (2010).
Therefore, Oklany et al. (1998) emphasize that due to its highly endothermic nature and equilibrium limitations, high temperatures are required to achieve a high methane conversion.

\[
CH_4(g) + H_2O(g) \rightarrow 3H_2(g) + CO(g)
\]

(1)

Traditional techniques such as coagulation, absorption in activated carbon, reserve osmosis and others can usually be used to remove these pollutants; however, these methods are not destructive and have a high price to remove these pollutants post-treatment (Aranha et al., 2023).

In this context, Liu et al., (2009) state that not only does the SMR process occur, but there is also the presence of other distinct catalyzed reactions, such as the water-gas shift (WGS), a reversible reaction (equation 2) in which CO reacts with H₂O in the forward direction, producing H₂ and CO₂. In order to eliminate the produced CO, Low Temperature Shift (LTS) and High Temperature Shift (HTS) reactions are employed, representing reactions occurring at temperatures of 200°C and 400°C, respectively (Zanoteli et al., 2014).

\[
CO(g) + H_2O(g) \leftrightarrow H_2(g) + CO_2(g)
\]

(2)

2.2 Chemical Kinetics

A catalyst is a substance responsible for increasing the rate of a reaction and can operate in various ways, such as modifying the force between reactants, blocking side reactions, or facilitating the breaking of chemical bonds (Evelio & Acevedo, 2006). Regarding the SMR process, from an industrial perspective, nickel is considered the most suitable catalyst for this process due to factors of availability and low cost. However, the deposition of carbon on active sites directly affects the activity of these catalysts (Barroso-Quiroga & Castro-Luna, 2010).

According to Acevedo (2006), when a chemical reaction is catalytic, it occurs on the layer of reactants and products that have been adsorbed on the catalyst's surface. Thus, there are several kinetic models responsible for describing the mechanism and kinetic data of this type of reaction. For the simulation of reactions, the choice of the kinetic model is necessary as it becomes useful for reactor design, as well as for control and optimization purposes (Amran et al., 2017).

In this work, the chosen model was the Langmuir-Hinshelwood-Hougen-Watson (LHHW) model, the same model used in the study conducted by Amran et al. (2017) for the simulation of SMR and WGS reactions. The authors denote that the expression for the reaction rate according to this model is given by the following general equation:

\[
r = \frac{\text{kinetic factor} \cdot \text{driving force}}{\text{adsorption term}}
\]

(3)

In this work, the reaction rates (4) and (5) were used for the SMR and WGS reactions, respectively:

\[
-r_1 = \frac{k_1P_{CH_4}}{1 + K_{H_2}P_{H_2}}
\]

(4)

\[
-r_2 = \frac{k_2P_{CO}P_{H_2} - k'_2P_{H_2}P_{CO_2}}{1 + K_{H_2}P_{H_2}}
\]

(5)
Where,

\[ k_i = A_i e^{-\frac{E_{a_i}}{RT}} \]

\[ A_1 = 6.6204 \times 10^6, \quad A_2 = 6.5544 \times 10^7, \quad A_2' = 8.8019 \times 10^9, \quad E_{a_1} = 88490 \text{ KJ/mol}, \quad E_{a_2} = 120000 \text{ KJ/mol}, \quad E_{a_2'} = 160738.6 \text{ KJ/mol}, \quad \text{and } K_{H_2} = 4.0530 \] are constants. The constant K that may appear in the driving force and adsorption terms are defined in Aspen Plus according to the equation, as they may be temperature-dependent (ASPEN PLUS HELP, 2017).

\[ \ln(K) = A + \frac{B}{T} + C \ln(T) + D T \]  

(6)

Where,

T is the temperature, A, B, C and D constants. Thus, in Aspen Plus, \( \ln(K_{H_2}) = 1.3996 \).

2.3 Industrial Chemical Reactors

Depending on the process, there may be a need for the use of continuous tubular reactors, whose generalized model is provided by the plug flow reactor (PFR) in Figure 1. Missen et al. (1999) assert that this type of reactor consists of an empty tube or contains some filling, such as catalyst particles, and can be used for both liquid-phase and gas-phase reactions, whether on a laboratory scale or a large scale. The PFR, according to Perry et al. (1997), is primarily suitable for cases that require significant heat transfer, high pressures, very high or very low temperatures, and for situations where relatively short reaction times are satisfactory.

![Figure 1: Representation of the PFR](Source: Levenspiel, 1999)

2.4 Thermodynamic Method (Property Method)

A property method can be understood as a set of methods and models that Aspen Plus uses for the determination of thermodynamic and transport properties (ASPEN PLUS HELP, 2017). In this work, the Peng-Robinson (PENG-ROB) method was employed. This method is recommended for applications involving gas processing, refining, and petrochemicals, where nonpolar or slightly polar mixtures predominate, and it is suitable for regions with high temperature and high pressure (ASPEN PLUS HELP, 2017).
2.5 Latin Hypercube Sampling (LHS) Sampling Plan

The first step in constructing and evaluating a surrogate model in the simulation and modeling of processes is the development of a sampling plan. The correct evaluation of a model or the response variables of a project depends, among other factors, on the choice of a good sampling plan, i.e., the selection of decision variables that can generalize well and allow for a good visualization of the behavior being assessed (Forrester et al., 2008).

An example of this is the Latin square, a sampling plan for discrete variables in two dimensions. To create n cases, a Latin square of order n is formed by n rows and n columns, filled with characters belonging to the set \{1, 2, ..., n\}, so that each character appears only once in each row and column (Forrester, et al., 2008). For example, a representation of a Latin square of order 5 can be seen in Figure 2.

The LHS can be extended to an n-dimensional setting as described above. According to Santner et al. (2018), to obtain an LHS of n points (cases) and k dimensions, the standard approach involves dividing each unitary axis (dimension) \([0,1]\) into n subintervals of equal size, such as \([0, 1/n), (1/n, 2/n), ..., ((n – 1) / n, 1]\).

This division implies partitioning the design space into \(n^k\) cells of equal size. Then, one of the characters from the set \{1, 2, ..., n\} is randomly chosen. In each of the cells containing this character, a value is randomly selected for each dimension within the respective interval in which the cell is located, resulting in the n resulting points (Santner et al., 2018). Figure 3 illustrates an LHS of 8 points and 2 dimensions.

![Latin Square of Order 4](Figure 2: Latin Square of Order 4)
Source: FORRESTER et al., 2008

![LHS of 8 Points and 2 Dimensions](Figure 3: LHS of 8 Points and 2 Dimensions)
Source: SANTNER et al., 2018
2.6 Cross-Validation

Cross-validation or K-fold method is a technique applied to evaluate the generalization capacity of a model through a simulated dataset. This method selects the best model from a set of models. The cross-validation method randomly divides the total amount of data into groups of relatively equal sizes, where in each iteration, one of the groups is removed and referred to as the validation set. The mean square prediction error (equation 7) is calculated for the removed group, respectively.

\[ MSPE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]  

(7)

The process will repeat until the last group, and at the end, an average of the Mean Square Prediction Errors (MSPE) from each of the groups will be calculated.

2.7 Linear Regression

Linear regression is a method that fits a line through data. We can define it as simple linear regression and multiple linear regression. Simple linear regression is described as the linear relationship between a dependent variable and an independent variable.

\[ Y = \beta_0 + \beta_1 X_i + \epsilon_i \]  

(8)

Where,

\( Y \) is the response variable, \( \beta_0 \) and \( \beta_1 \) are parameters, \( X \) is the independent variable, and \( \epsilon \) is the measurement error.

Multiple linear regression, on the other hand, establishes an equation that will be used to predict the values that \( Y \) will take into account multiple independent variables.

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_n X_{in} + \epsilon_i \]  

(9)

2.7.1 Quadratic linear regression and pure quadratic

Quadratic linear regression is a model or a particular case of multiple linear regression that allows you to obtain the equation of a parabola. This model is studied through multiple linear regression methods. Pure quadratic, on the other hand, is a constant expression that includes linear terms and those that are purely quadratic in the predictors. The model or equation that defines quadratic regression is:

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i \]  

(10)

2.7.2 Linear regression with interaction

Linear regression with interaction is another type of multiple linear regression. This type of regression is more complex because, in the other models, the change between them was the intercept of the variables, whereas in this model, a new term called interaction is included, which allows the slope of the line to change from one group to another. If we adjust the model with this equation, we have:
In this case we have $\alpha_1 ZX$ as an interaction term.

### 2.8 Gaussian Process (Kriging)

The study of Gaussian random processes is largely a study of covariance functions or correlation functions (Clelto Barros Gomes, 2009). A Gaussian process can also be described as an uncountably infinite set of random variables, with the property that any finite subset of such variables has a multivariate normal distribution of dimension $p$, a mean vector $\mu$, and a covariance matrix $\Sigma$. In this process, the distance between two points is referred to as the kernel, understood as a function of two points in space.

AI tools like machine learning and kriging processes can benefit and span across various sectors. Through the strategic application of AI-driven technologies such as machine learning, natural language processing (NLP), and computer vision, significant advancements can be achieved. (Almustafa et al., 2023)

#### 2.8.1 Exponential quadratic and exponential kernel

This kernel is understood as a Gaussian function where $x$ is the maximum point, and the distance increases as we move away from that point (12). The exponential kernel is identical to the exponential quadratic, except for the Euclidean distance, which is not squared (13).

$$K(x, x') = \sigma^2 \exp \left( -\frac{(x - x')^2}{2l^2} \right)$$ (12)

$$K(x, x') = \sigma^2 \exp \left( -\frac{r}{\sigma_f} \right)$$ (13)

Where, $r = \sqrt{(x - x')^T (x - x')}$ (14)

#### 2.8.2 Matern 5/2 kernel

The Matern kernel has a parameter $\nu$ that controls the smoothness of the resulting function. For Matern 5/2, the value of $\nu = \frac{5}{2}$ is one of the most common because it is not infinitely differentiable.

$$K(x, | \theta) = \frac{1}{\Gamma(\nu)} \frac{2^{\nu - 1}}{\sigma^\nu} \left( \frac{\sqrt{2\nu} D}{\sigma} \right)^\nu K_\nu \left( \frac{2D}{\sigma} \right)$$ (15)

$$K(x, x' | \theta) = \sigma_f^2 \left( 1 + \frac{\sqrt{5D}}{\sigma_l} + \frac{5D^2}{3\sigma_l^3} \right) \exp \left( \frac{\sqrt{2D}}{\sigma_l} \right)$$ (16)
2.8.3 Rational square kernel

The rational square kernel can be characterized by a mixture of scales of exponential quadratic kernels with different specific length scales. The equation representing the rational square is:

\[ K(x, x'|\theta) = \sigma_f^2 \left( 1 + \frac{D^2}{2\alpha \sigma_f^2} \right) \]  

(17)

2.9 RMSE – Root Mean Square Error

The RMSE (Root Mean Square Error) is a measure responsible for calculating the square root of the mean of the errors between true values and predicted values. This indicator has a proportion equal to the dimension of true and predicted values, interpreted as the average deviation between the true and predicted values.

\[ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \]  

(18)

A lower RMSE value indicates a more satisfactory fit because it represents the model's ability to predict the response.

2.10 R\(^2\) - R-Squared

R\(^2\) represents the variance ratio for a dependent variable that is explained by independent variables through a regression model.

\[ R^2 = 1 - \frac{\sum(y_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \]  

(19)

The R\(^2\) value should be between 0 and 1, with values closer to 1 indicating better agreement of the model.

3 METHODOLOGY

3.1 Base Case

For the simulation construction, Aspen Plus version 10, a commercial software, was used. Initially, in the Components folder of the Properties section, the components present in reactions (1) and (2) were defined, as shown in Figure 4. Then, in the Methods folder of the same section, the base thermodynamic method was chosen, as shown in Figure 5.
Use of Machine Learning Techniques in the Modeling of an Industrial Reaction System

Figure 4: Components present in the reactions.
Source: Author's own work, 2021

Figure 5: Determination of the thermodynamic method.
Source: Author's own work, 2021

Subsequently, in the Simulation section, the following specifications were made. The process flowchart was constructed by choosing an RPlug reactor, a rigorous model of a PFR with reactions controlled by kinetics, containing a feed stream and a product stream.

In the Streams folder, the input variables related to the feed stream were specified, including temperature, pressure, total flow rate, and composition. The basis for total flow rate and composition was also selected. Then, in the Blocks folder, the type of PFR was specified, and the reactor's configurations regarding its dimensions and valid phases were defined.

The creation of reactions was performed in the Reactions folder. First, it was necessary to define a name for identification, as well as the kinetic model that describes these reactions. In the next step, the stoichiometries of the reactions were defined. Figure 6 demonstrates the definition of stoichiometry and the type of reaction (kinetics) for the reaction described in (1). Similarly, this was done for the direct and reverse reactions described in Figure 7 shows the defined reactions.
When introducing the kinetics of the reactions, it was necessary to specify the rate for each reaction (equations 4 and 5), according to equation (3). For equation (4), the kinetic factor term was defined with parameters such as the pre-exponential factor, temperature exponent, and activation energy. Similarly, the kinetic factor was defined for equation (5) based on the direct and reverse reactions (2). Regarding the driving force expression, in Aspen Plus, it needs to be specified in two terms representing the direct reaction (term 1) and the reverse reaction (term 2).

Thus, for equation (4), in term 1, only the methane concentration exponent was defined, as methane is the only component present in the driving force term, and since the constant multiplying this concentration is equal to 1 (ln(1) = 0 in Aspen Plus), the coefficient was defined as 0. Since there is no reverse reaction in (1), term 2 is null, and therefore, 0 is introduced for the concentration exponents and a high negative number for the constant to make it approximately 0 ($e^{-1\times10^{15}}$). Figure 8 demonstrates these specifications. Similarly, the driving force term was defined for equation (5) with a difference in term 1 for the concentration exponents, which were set to 1 for the reactants of the direct and reverse reactions (2).
Regarding the adsorption expression, the exponent was initially defined. It can be observed that it consists of two terms, the first formed by a constant and the second containing the multiplication between the adsorption constant and the methane concentration. Thus, the concentration exponent and the adsorption constant for the first term were specified as 0 since the first term must be equal to 1. For the second term, the concentration exponent was defined as 1, and the adsorption constant $\ln(K_{H_2})$ was specified. This same specification was repeated for the three reactions.

### 3.2 Sensitivity Analysis

The creation of sensitivity analysis was carried out in the Model Analysis Tools/Sensitivity folder. In the Vary tab, the variables to be manipulated were defined, which include the diameter and length of the reactor, as well as the temperature and pressure of the feed stream (Figure 9). In the Define tab, the controlled variables were defined, including the molar flow rates of all components in the product stream and the temperature in the product stream.

The molar flow rates of $CH_4$, $H_2O$ and $CO$, in the feed stream were also defined to enable the definition of conversions in the Tabulate tab, which are also controlled variables (Figures 10 and 11). The selected disturbance variables included the molar flow rates of $CO$, $H_2$, $CO_2$ and $N_2$, in the feed stream and the molar flow rate of the feed stream.
For the construction of cases, a percentage variation of +/- 20% relative to the base case values was considered for the manipulated variables, and a Matlab function was used to calculate the Latin Hypercube Sampling (LHS) with 200 cases. Thus, the generated cases were included in the Case Study mode.

3.3 Metamodels

analyses of metamodels were performed using Matlab 2019a software. A table of size 200 x 15 was obtained, containing the dependent and independent variables used as a reference in building the metamodels. The Regression Learner, a model used for predictions using supervised machine learning in Matlab, was used to find the correlation between the dependent and independent variables. The cross-validation method was used for each independent variable with respect to the selected dependent variable for analysis.

The types of models used for dependent variables from 6 to 15 within the Regression Learner included linear regression and Gaussian process regression (Kriging). For linear regression, advanced linear regression options known as linear (8), interaction (11), quadratic and pure quadratic (10) were applied to the variables. For Gaussian process regression (Kriging), kernels such as rational square (17), exponential (13), exponential quadratic (12), and Matern 5/2 (16) were applied. To determine which metamodel best suits the case, RMSE (Root Mean Square Error) (18) and $R^2$ (R-Squared) (19) measures were used to assess the accuracy of the
fit of the points in graphs for each variable. These measures were obtained after applying each of the models to the points obtained from the variables.

4. RESULTS AND DISCUSSION

4.1 Base Case

After running the base case, the results converged. Figures 12 and 13 provide the results obtained for the reactor outlet stream and the mass and energy balances in the reactor, respectively. It can be observed that there is a decrease in the molar flow rate of \( CO(g) \), which was expected since the goal of the WGS reaction was not only to produce \( H_2(g) \), but also to reduce the amount of \( CO(g) \) present in the reactor. An expected decrease in the molar flow rates of \( CH_4(g) \) and \( H_2O(g) \) the reactants and was also observed, along with an expected increase in the molar flow rates of the products \( H_2(g) \) and \( CO_2(g) \), as well as a constant flow rate of \( N_2(g) \).

<table>
<thead>
<tr>
<th>Mole Flows</th>
<th>Units</th>
<th>FEED</th>
<th>PRODUCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH4</td>
<td>kmol/hr</td>
<td>7596</td>
<td>8271,39</td>
</tr>
<tr>
<td>H2O</td>
<td>kmol/hr</td>
<td>1454,43</td>
<td>1116,73</td>
</tr>
<tr>
<td>CO</td>
<td>kmol/hr</td>
<td>2199,16</td>
<td>669,246</td>
</tr>
<tr>
<td>H2</td>
<td>kmol/hr</td>
<td>882,124</td>
<td>47,596</td>
</tr>
<tr>
<td>CO2</td>
<td>kmol/hr</td>
<td>1313,73</td>
<td>3499,06</td>
</tr>
<tr>
<td>N2</td>
<td>kmol/hr</td>
<td>702,245</td>
<td>1874,57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mole Fractions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CH4</td>
<td>0.191473</td>
</tr>
<tr>
<td>H2O</td>
<td>0.289516</td>
</tr>
<tr>
<td>CO</td>
<td>0.110113</td>
</tr>
<tr>
<td>H2</td>
<td>0.172953</td>
</tr>
<tr>
<td>CO2</td>
<td>0.092625</td>
</tr>
<tr>
<td>N2</td>
<td>0.157465</td>
</tr>
</tbody>
</table>

Figure 12: Results obtained for the product stream.
Source: Author’s own work, 2021

<table>
<thead>
<tr>
<th>Total</th>
<th>Units</th>
<th>In</th>
<th>Out</th>
<th>Generated</th>
<th>Relative difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mole-flow</td>
<td>kmol/hr</td>
<td>7596</td>
<td>8271,39</td>
<td>675,39</td>
<td>0</td>
</tr>
<tr>
<td>Mass-flow</td>
<td>kg/hr</td>
<td>150470</td>
<td>150470</td>
<td>1.12145e-12</td>
<td></td>
</tr>
<tr>
<td>Enthalpy</td>
<td>kcal/hr</td>
<td>-2,25317e+08</td>
<td>-2,25317e+08</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13: Mass and energy balances in the reactor.
Source: Author’s own work, 2021

Figure 14 and 15 shows the variation in the molar fractions of and along the length of the reactor, indicating that the majority of the reactions occur during the first meter of reactor length. Additionally, due to the highly endothermic nature of the SMR reaction and the use of an adiabatic reactor, there is a decrease in temperature along the length of the reactor (Figure 16).
Use of Machine Learning Techniques in the Modeling of an Industrial Reaction System

Figure 14: Molar fraction of \( CO_{(g)} \) along the length of the reactor
Source: Author's own work, 2021

Figure 15: Molar fraction of \( H_2{(g)} \) along the length of the reactor.
Source: Author's own work, 2021

Figure 16: Temperature variation along the length of the reactor.
Source: Author's own work, 2021
4.2 Sensitivity Analysis

By analyzing the metamodels, it was possible to collect the RMSE and $R^2$ results for each of the manipulated variables. This analysis also helped in determining which metamodel fits best for each of the situations presented by the variables. The results of the process for linear regression and Gaussian process are presented in Tables 1 and 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>RMSE</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.0036988</td>
<td>1.00</td>
<td>0.003613</td>
<td>1.00</td>
<td>0.0036218</td>
<td>1.00</td>
<td>0.0035355</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>0.0022631</td>
<td>0.97</td>
<td>0.0020781</td>
<td>0.97</td>
<td>0.0013496</td>
<td>0.99</td>
<td>0.0007921</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>0.0049786</td>
<td>1.00</td>
<td>0.0048524</td>
<td>1.00</td>
<td>0.0046261</td>
<td>1.00</td>
<td>0.0044976</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>10.504</td>
<td>1.00</td>
<td>5.8997</td>
<td>1.00</td>
<td>10.561</td>
<td>1.00</td>
<td>5.9297</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>21.935</td>
<td>1.00</td>
<td>10.978</td>
<td>1.00</td>
<td>19.812</td>
<td>0.99</td>
<td>10.067</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>1.7463</td>
<td>0.98</td>
<td>1.9656</td>
<td>0.98</td>
<td>0.72358</td>
<td>1.00</td>
<td>0.24616</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>41.679</td>
<td>0.99</td>
<td>21.589</td>
<td>1.00</td>
<td>4.7515</td>
<td>1.00</td>
<td>0.96768</td>
<td>0.99</td>
</tr>
<tr>
<td>13</td>
<td>9.1775</td>
<td>1.00</td>
<td>4.7515</td>
<td>1.00</td>
<td>0.72358</td>
<td>1.00</td>
<td>0.24616</td>
<td>1.00</td>
</tr>
<tr>
<td>14</td>
<td>2.3159x10^-6</td>
<td>1.00</td>
<td>2.3159x10^-6</td>
<td>1.00</td>
<td>2.4193x10^-6</td>
<td>1.00</td>
<td>2.4193x10^-6</td>
<td>1.00</td>
</tr>
<tr>
<td>15</td>
<td>0.41705</td>
<td>1.00</td>
<td>0.24616</td>
<td>1.00</td>
<td>0.62573</td>
<td>0.99</td>
<td>0.62573</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Source: Author's own work

Table 1 - Results for Linear Regression

<table>
<thead>
<tr>
<th>Output variables</th>
<th>Linear Regression</th>
<th>Gaussian Process (Kreissing)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>$R^2$</td>
</tr>
<tr>
<td>6</td>
<td>0.0018487</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>0.0007491</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>0.0023994</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>3.0757</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>4.9476</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>0.56473</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>10.66</td>
<td>1.00</td>
</tr>
<tr>
<td>13</td>
<td>2.2171</td>
<td>1.00</td>
</tr>
<tr>
<td>14</td>
<td>0.00056675</td>
<td>1.00</td>
</tr>
<tr>
<td>15</td>
<td>0.23178</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Source: Author's own work

Through the table, it was possible to notice that the models with better values are those with lower RMSE values and $R^2$ equal to 1, proving their higher precision. So, for each dependent variable, we have a regression model that fits best.

In the linear regression process, it was observed that the equation model that best suited the variables Conversion of CH$_4$ (mol) (6), Conversion of CO (mol) (7), Conversion of H$_2$O (mol) (8), Water Outlet Flow Rate Ratio (kmol/h) (10), Carbon Monoxide Outlet Flow Rate Ratio (kmol/h) (11), Hydrogen Outlet Flow Rate Ratio (kmol/h) (12), Carbon Dioxide Outlet Flow Rate Ratio (kmol/h) (13), and Outlet Temperature (ºC) (15), was the quadratic model (Figure 17 and 18). For the variable Methane Outlet Flow Rate Ratio (kmol/h) (9), the most effective model was the interaction model (Figure 19), and for Nitrogen Outlet Flow Rate Ratio (kmol/h) (14), the most effective model was linear (Figure 20).
In the Gaussian process, variables 6 and 15 adapted to two models, the rational quadratic and exponential quadratic (Figure 19). Variables 7, 11, 12, and 14 adapted to the Matern 5/2 kernel (Figure 20). Variables 13 and 15 behaved better with the rational quadratic, and finally, variables 9 and 10, where the regression was better adapted to the exponential quadratic kernel.

Figure 17 - Quadratic model for Methane Conversion.
Source: Author's own work, 2021

Figure 18 Quadratic model for Hydrogen Outlet Flow Rate Ratio.
Source: Author's own work, 2021
Use of Machine Learning Techniques in the Modeling of an Industrial Reaction System

Figure 19 - Interaction model for Methane Outlet Flow Rate Ratio
Source: Author's own work, 2021

Figure 20 - Representation of the linear model for Nitrogen Outlet Flow Rate Ratio.
Source: Author's own work, 2021
5 CONCLUSION

It was possible to understand the importance of studying hydrogen production, including the reactions and kinetics involved in the process, as well as its applications. Regarding the developed process, the purpose of hydrogen generation and the reduction of carbon monoxide.
Use of Machine Learning Techniques in the Modeling of an Industrial Reaction System

...in the reactor outlet were observed. Furthermore, the influence of the SMR reaction on the temperature variation along the length of the reactor was evident.

Moreover, it can be concluded that through the sensitivity analysis of each variable, by means of RMSE and R², both linear and Gaussian process regressions presented models that were better adapted to most of the independent variables. For example, the quadratic model in linear regression and the Matern 5/2, Rational Quadratic, and Exponential Quadratic models in Gaussian process regression proved to be effective for various independent variables.

REFERENCES


CARVALHO, A. X. Y.; GOÉS, G. S. Introdução ao Software R e à Análise Econometrícia - Seleção de Variáveis em Modelos de Regressão.


GOMES, José Clelto Barros. Estimação não paramétrica para função de covariância de processos Gaussianos espaciais/ José Clelto Barros Gomes -- Campinas, [S.P.:s.n],2009.


