# Hydrogen Production Prediction Using XGBOOST: A Comparative Study of Energy Conversion Scenarios


## Authors

Diosael Ferreira Sacramento[1], Tainan Sousa Viana[2], Paulo Alexandre Rocha[3]

[1]Federal University of Ceará – Pici Campus , e-mail: diosaelferreira@alu.ufc.br , Fortaleza (Brazil).

[2] Federal University of Ceará – Pici Campus , e-mail: tainanviana@alu.ufc.br, Fortaleza (Brazil).

[3]Federal University of Ceará – Pici Campus , Advisor , e-mail: paulo.rocha@ufc.br, Fortaleza (Brazil).



## ABSTRACT

This study evaluates the performance of the XGBoost algorithm in predicting green hydrogen production from wind energy, comparing five energy conversion scenarios with different physical and empirical models. Time series of wind speed data from the state of Ceará (2019–2021) were used, subjected to cleaning and seasonal imputation, with 80% of the data divided for training and 20% for testing. The results indicated distinct performances among the test scenarios: C1 (RMSE = 201.95 kg/km²; $R^2$ = 0.72), C2 (RMSE = 86.33 kg/km²; $R^2$ = 0.54), C3 (RMSE = 280.53 kg/km²; $R^2$ = 0.73), C4 (RMSE = 45.19 kg/km²; $R^2$ = 0.49), and C5 (RMSE = 220.74 kg/km²; $R^2$ = 0.72). Scenario C4 presented the best Skill Score (0.83) and the lowest mean error (MAE = 31.72 kg/km²), standing out for its physical accuracy and stability of predictions. C2 demonstrated the shortest processing time (649 s) and good relative accuracy (MAPE = 24.02%). The results confirm that integrating physical modeling and machine learning enhances accuracy and computational efficiency.



***Keywords**: Green Hydrogen Forecasting, Machine Learning, XGBoost , and Wind Energy.*


## 1. INTRODUCTION

Most countries in the world still rely heavily on fossil fuels to meet their energy needs. However, the use of these fuels causes many negative environmental consequences. Hydrogen has emerged as a promising alternative to achieve an environmentally friendly future that reduces carbon emissions. This characteristic highlights why accurate forecasting of renewable production is crucial to optimizing the production profile and overall cost of green hydrogen.

Carbon emissions can be reduced through the use of renewable energy sources, including solar energy, wind energy, biogas, and alternative green fuels. Hydrogen has emerged as a promising alternative to achieve an environmentally friendly future that reduces carbon emissions. The current fascination with this specific source stems from its environmentally friendly characteristics and efficient combustion properties ( Allal et al., 2024).

Weather forecasting consists of a method for estimating a future value using historical data. Several methods exist for this simulation; the authors Kumar, Pal , and Tripathi (2020) separated forecasting into models based on time scale, historical/meteorological observation data, and physical and mathematical techniques/models. Other researchers have classified the methods as numerical weather prediction, traditional statistical methods, Artificial Intelligence (AI), and hybrid methods (LIU; TIAN; LI, 2012).

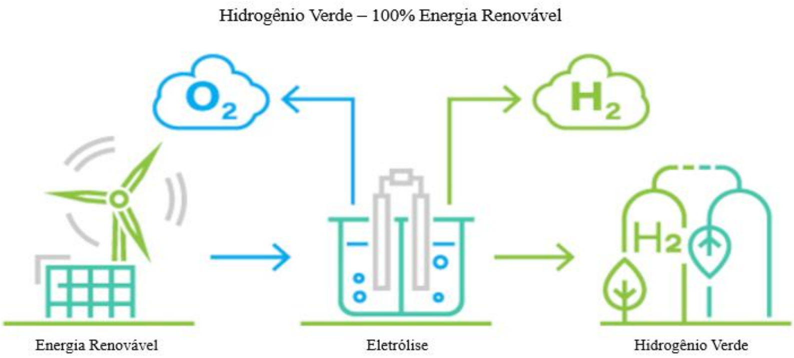
This study aims to develop and compare conversion scenarios and machine learning techniques to estimate green hydrogen production from wind power potential in the regions of Ceará.


10.18086/swc.2025.05.12 Available at http://proceedings.ises.org

## 2. LITERATURE REVIEW

The production of green hydrogen occurs through a chemical reaction called electrolysis, which is the breaking down of a hydrogen-containing compound molecule, such as water, to obtain hydrogen. In this case, the electrolysis process is initiated by electricity generated from renewable sources such as wind or solar power. For this reason, the hydrogen obtained is called green: none of its production processes result in the emission of polluting gases into the atmosphere, nor does it release substances capable of degrading the environment. At the end of electrolysis, in addition to hydrogen, oxygen is released ( Guitarrara , 2025), as shown in Figure 1.

Figure 1: Hydrogen production process via water electrolysis.




Source: Adapted from Guitarrara (2025).

A wind turbine is used to convert mechanical energy into electrical energy ; the electrical power generated by a wind turbine varies with the cube of the wind speed across the area swept by the turbine blades. Therefore, techniques for converting wind to hydrogen can be applied using the Wind Power Law, which can be calculated as follows.

$$P(v) = \frac{1}{2}\text{Cp,eq}\ \rho A\ V^3 \tag{1}$$

Where, P is Wind Power (kg/m), ρ equals 1.225 kg/m³ (air density), A (pi*r^2, area swept by the turbine (m^2)), Cp,eq are the constants equivalent to the power coefficient, v is the wind speed (m/s). Another relationship used to estimate the power output of a wind turbine is the capacity factor or turbine efficiency, which indicates how much a generating (wind) turbine actually produces in relation to its maximum theoretical potential . This relationship can be seen in Figure 2, where the power increases up to nominal speed, so the turbine will operate constantly at the wind turbine's nominal power.

$$Fc = \frac{P(v)}{Pnom} \tag{2}$$

Where P(v) is the power generated through wind speed and Pnom is the nominal power of the wind turbine. From this value of P(v), in a second stage, Equation (2.3) is one of the methods used to estimate hydrogen production ( Alavi et al., 2016; Mostafaeirpour et al., 2019):

$$H2 = \frac{n \times P(v)}{Econs} \tag{3}$$

Where H2 is hydrogen production, on is the efficiency of the electrolysis process, and Econs is the electricity consumption of the electrolyzer . The power curve of a wind turbine in Figure 2 is divided into three main operating regions (lower cutoff, operating region, and upper cutoff) ( Manwell et al., 2009), characterized by a cubic function between the lower and nominal cutoff speeds and a constant power up to the upper cutoff speed.

Figure 2: Power curve of a wind turbine


Source: ( Manwell et al., 2009).

As indicated by the power curve in Figure 2, the conversion of wind energy into electrical energy occurs from a minimum speed called the starting speed ( vp ), usually values of the order of 2.5 - 3 m/s; below these values, the energy content of the wind does not justify its use (AMARANTE et al, 2001). The turbine power is limited to the nominal value ( Pn ) at the nominal speed ( vn ), typically speeds between 12 and 15 m/s. For speeds between vp and vn, the electrical power generated by the wind turbine increases with the cube of the speed according to equation (1). From the nominal speed, the turbine speed control maintains the generated power constant until, for speeds exceeding the cut-off speed ( vc ), generally 25 m/s, the turbine is taken out of operation in order to protect it from excessive mechanical stress (CUSTÓDIO, 2009).

## 3. METHODOLOGY

First, data was collected, followed by an exploratory analysis of the time series. The XGBoost algorithm was used to run the models. The time series was divided into two subsets: one for training and the other for testing. The training subset encompassed 80% of the total data (42,096 records, from 01/01/2019 to 24/07/2021), while the remaining 20% (10,512 records) were used as the test set, covering the period from 25/07/2021 to 31/12/2021. The model evaluation was conducted using a comprehensive set of classical and normalized statistical metrics for point forecasting.

### 3.2.1 Conversion method used by Javaid et al (2022)

Scenario 1 or (C1), used in Javaid et al. (2022), assumes constant behavior for the efficiency of an electrolyzer , adopting a simplified approach in which the electrolyzer efficiency is treated as constant throughout the entire operating period. This methodology assumes that the system operates in steady state, allowing the conversion equation to be simplified to a direct relationship between the supplied electrical energy and hydrogen production. In Yilmaz et al. (2021), for example, hydrogen production in a hybrid solar-wind system was estimated considering a fixed efficiency of 70%. The study sought to identify the potential hydrogen generation capacity in different regions of Turkey, using hourly data on solar irradiance and wind speed. Similarly, Nayak et al. (2023) evaluated hydrogen production in off-grid systems in the Indian context, also assuming constant efficiency ($\eta = 75\%$). In this case, the objective was to compare the levelized cost of hydrogen (LCOH) under different turbine sizes and electrolyzer capacities . This work shows that, although the fixed-efficiency model simplifies the simulation, it ignores the dynamic behavior of the electrolyzer at partial loads and the transient losses associated with the intermittency of wind power.

### 3.2.2 Methodology presented by Zhu et al (2022)

Scenario 2 (C2) represents efficiency as discrete operating intervals, where each power range corresponds to an average efficiency. This type of modeling was originally proposed by Zhu et al. (2022), who developed an empirical alkaline electrolyzer model from experimental data. The authors collected measurements of current, voltage, and hydrogen production under different power levels, grouping the results into five relative load ranges. For each range, they calculated the average efficiency based on the ratio between useful energy and electrical energy consumed.

The model was applied to a hybrid wind-solar system, and the results showed that the overall efficiency of the electrolyzer varied between 72% and 82% depending on the load factor, revealing a downward trend at higher loads. Similarly, Ould Amrouche et al. (2020) applied the same methodology to an autonomous wind system in Algeria, considering efficiency ranges to estimate hydrogen production and assess the system's energy stability. Caglayan et al. (2022), in a study for Northern Europe, implemented a model based on discrete efficiency ranges on a continental scale, using the REMix software to estimate the contribution of green hydrogen to the European energy matrix until 2050.

Initially, the same power calculated in scenario 1 was considered, only varying the electrolyzer efficiencies in turbine load ranges as per table 1. This value indicates the percentage of the electrolyzer 's nominal power that is... being used . The power supplied to the electrolyzer is compared to the equipment's nominal power to determine the efficiency range in which the system is operating . The ratio between the generated power and the electrolyzer 's nominal power ( pot_electrolyzer ) is calculated using the following equation:

$$faixa_{potência} = \frac{num\_aero \times P}{Pot\_eletrolisador} \tag{4}$$

Electrolyzer Power Range and Efficiency ( n)

| **power_range** | **Efficiency ( n )** |
|---|---|
| x < 0.05 | 0% (Off) |
| 0.05 ≤ x < 0.20 & x < 0.2 | 82% |
| 0.2 ≤ x < 0.5 & x < 0.5 | 77% |
| x0.5 ≤ x < 1.0 & x < 1.0 | 73% |
| 1.0 ≤ x < 1.5 & x < 1.5 | 66% |

It is observed that the system remains off when the input power is less than 5% of the nominal power, that is, at energy levels insufficient to initiate electrolysis. In the low load range between 5% and 20%, the efficiency reaches its maximum value of 82%, showing that the electrolyzer operates more efficiently under partial load conditions, with less heat dissipation and lower ohmic losses. As the load increases, the efficiency shows a tendency to gradually decrease, dropping to 77%. and 73%.

In the overload range, a significant reduction in efficiency to 66% is observed, resulting from increased internal resistance, elevated temperature, and higher energy consumption per unit of hydrogen produced. This drop also reflects the actual behavior of electrolyzers , which exhibit decreasing efficiency when operating above nominal power, as described in Figure 4 of Section 1. Therefore, Table 4 shows that the optimal performance of the electrolyzer occurs around 10% to 20% of its nominal power, while continuous operation at high loads compromises the overall efficiency of the system. This relationship is essential for hydrogen production calculations, as it allows for a more realistic estimation of efficiency variations over time, especially under fluctuating wind conditions.

### 3.2.3 Methodology presented by Hofrichter et al. (2023)

Scenario 3 or (C3) is based on the modeling proposed by Hofrichter et al. (2023), identical to method 2, but proposes a continuous efficiency curve based on a polynomial function of order 3 or higher, which expresses the electrolyzer efficiency as a function of the load factor where the turbine input power is divided by its nominal power ( $fc = \frac{Pin}{Pnom}$). This developed approach carried out experiments with alkaline electrolyzers under variable load conditions, collecting voltage, current, and hydrogen flow data for different power levels. From these data, they fitted a polynomial function of the type:

$$\eta eleT = a0 + a1fc + a2f^2c2 + a3f^3c3 \tag{5}$$

allowing for the accurate representation of smooth efficiency variations across the entire operating range. Other authors, such as Carmo et al. (2013), had already highlighted that the efficiency of PEM electrolyzers exhibits non-linear behavior, especially at intermediate loads, which reinforces the validity of the polynomial approach. Nikiforidis et al. (2020) also explored this behavior when modeling electrolyzers in hybrid systems to predict performance under intermittent operation, obtaining good correlation between the numerical model and experimental data. This methodology is particularly suitable for forecasting studies with high temporal resolution, as it eliminates discontinuities and better reflects the transient responses of the system.

Instead of using discrete efficiency ranges, as in Scenario 2 , this model incorporates polynomial curves to describe the relationship between applied power and electrolyzer efficiency , allowing for smoother electrolyzer efficiency. The model followed the same power parameters as the previous scenarios, only considering the power ranges (equation 4) to obtain the electrolyzer operating efficiency .

This value indicates the percentage of the electrolyzer 's nominal power that is being used . The efficiency ( n) is initialized to zero for all values and then adjusted based on intervals of the relative power range.

If power_range < 0.1, The efficiency is zero , indicating that the electrolyzer does not operate below 10% of its nominal power.

If 0.1 ≤ power_range < 0.15 & power_range < 0.15 **:** Efficiency is defined by a fifth-degree polynomial curve :

$$n = \frac{100 + (0.00005\text{x}^\wedge 5 - 0.0061\text{x}^\wedge 4 + 0.2372\text{x}^\wedge 3 - 4.2014\text{x}^\wedge 2 + 36.675\text{x} - 6287)}{100\text{n}} \quad (6)$$

If 0.15 ≤ power_range < 1.0 & power_range < 1.0 : Efficiency decreases linearly as the load increases.

$$n = \frac{(-0.149\text{x} + 74.977)}{100\text{n}} \quad (7)$$

This relationship suggests that efficiency gradually decreases as the power approaches the electrolyzer 's nominal capacity .

If power_range ≥ 1.0 power_range = 1.0: The electrolyzer 's efficiency reaches its maximum efficiency at nominal power.

### 3.2.4 Methodology presented by Jared Thomas et al. (2024)

Scenario C4 incorporates the actual power and efficiency curves of wind turbines, making the wind conversion stage closer to real physical conditions. This modeling considers the power coefficient Cp from the turbine efficiency curve, which relates the power extracted from the wind to the available kinetic power, as well as the lower, nominal, and upper cutoff speeds. Thomas et al. (2024) conducted a study on the direct coupling between wind turbines and electrolyzers , considering the actual curves of 1.5 and 3 MW turbines. The model includes the influence of aerodynamic losses, air density, and blade pitch control, generating a dynamic estimate of the useful power delivered to the electrolyzer . Furthermore, Bandi et al. (2019) demonstrated that the accuracy in estimating hydrogen production strongly depends on the power curve used, since small differences in Cp behavior can significantly alter the available energy. Sørensen (2021) reinforces this conclusion, showing that the use of generic turbine curves can generate errors of up to 15% in the annual energy estimate. Therefore, the C4 scenario is fundamental for studies that seek to integrate real wind turbine data and simulate the operation of hybrid systems with fidelity.

Unlike previous scenarios, where the power generated by the wind turbine was treated in a simplified or constant manner, here we incorporate the actual efficiency curve of the wind turbine as a function of wind speed , since the efficiency of the wind turbine is not constant , depending on the aerodynamics of the blades, the wind speed, and the losses inherent in the conversion of wind energy into electricity.

Power is calculated using equation 2.1 mentioned earlier, according to the power ranges of wind speeds, the power coefficient of the wind turbine, and in this case, the efficiency curve ( cp ). Wind turbines do not generate power for all wind speeds. Therefore, the *np.where* function is used to model different ranges:

v < 3.5 m/s → No power generated (wind below minimum operating speed).

3.5 m/s ≤ v < 12 m/s → Power follows the equation: P=0.5×cp×area×density×v^3.

12 m/s ≤ v < 25 m/s → Constant power (the wind turbine operates at nominal power).

v ≥ 25 m/s → No power generated (wind turbine shut down for safety reasons).

The efficiency of a wind turbine ( cp ) is defined as a polynomial function of wind speed . This allows for the representation of aerodynamic losses and non-linear effects, such as drag and variations in energy conversion. The equation for calculating wind turbine efficiency is:

ηaero = np.clip (−0.0003 xv^3 +0.02 + v^2 −0.4 xv + 2.5,0,1) (8)

Where np.clip (..., 0, 1) ensures that efficiency always remains between 0 and 1 .

This curve can be interpreted as follows:

very low or very high winds , efficiency drops to zero .

moderate wind speeds or between [insert values here ], the efficiency reaches maximum values close to 1 (100%). This model always attempts to maintain consistency in electricity generation for wind speeds within the operating region of the wind turbine.

### 3.2.5 Methodology presented by Zénó Farkas et al. (2011)

Scenario 5 or (c5) considers air density when introducing atmospheric corrections in the modeling of wind power and overall system efficiency. It is based on the principle that air density (ρ) varies with temperature and pressure. The model was originally developed by Farkas et al. (2011) for wind systems in mountainous regions of Hungary, demonstrating that temperature variations of 10 °C can alter the available wind power by up to 6%. Subsequently, Li et al. (2022) extended this analysis to predict hydrogen production in regions with high seasonal temperature variation, such as Inner Mongolia, incorporating atmospheric correction in the long-term forecast. Hasan et al. (2023) applied a similar methodology to offshore hybrid systems, considering both thermal variation and air humidity in density modeling.

These studies demonstrate that including local meteorological conditions is essential for accurate modeling, especially in tropical and subtropical countries where there is a large temperature range. Air density is calculated using the ideal gas law:

$$\rho = \frac{Pa}{R \cdot T} \tag{9}$$

Where, ρ and air density (kg/m³), atmospheric pressure ( Pa ) , R = 287.0 J/ kg·K is the gas constant for dry air and T is the air temperature (Kelvin). This adjustment is important because:

- At high altitudes or on hot days , air density decreases , reducing the power generated.
- At low temperatures or high pressure , density increases , thus increasing power generation.

The power extracted from the wind depends directly on the air density adjusted in equation 2.28 and is calculated by equation 1. Where, ρ is the adjusted air density, A is the area swept by the wind turbine blades, v is the wind speed, and cp is the power coefficient of the wind turbine. Here, we use the Betz limit , which establishes that the maximum theoretical efficiency of converting wind energy into electricity is 59.3% . The Betz limit establishes that at most 59.3% of the wind's kinetic energy can be converted into mechanical energy by a wind turbine; it is a theoretical , idealized limit and is never reached in practice . The limit is derived from physical principles: if a turbine extracted 100% of the wind's energy, the air would stop completely behind it, which is impossible because the airflow needs to keep moving. In realistic simulations or computational modeling, the Betz limit This serves as a superior theoretical reference , but a realistic turbine power coefficient (denoted as Cp ) should be used . The electricity generated by the wind turbine is converted into hydrogen by the electrolyzer , and the electrolyzer efficiency is considered fixed at 75% .

### 3.2.6 Synthesis of the Methods Used for Modeling Electrolyzers

Considering the information presented, the wind turbines used for calculating wind power generation were selected so that the hub height was compatible with the measurement height of the wind data. The main technical information of the wind turbines is shown in Table 2.

**Table 2 –** Technical information of the selected wind turbines

| **Information** | **AW125/3000** |
|---|---|
| Manufacturer | Acciona |
| Rated power | 3,000 kW |
| Lower cutting speed | 3.5 m/s |
| Nominal speed | 12 m/s |
| Superior cutting speed | 25 m/s |
| Rotor scan area | 12,305 m² |
| Hub height | 120 m |
| Application | CE |

Source: Prepared by the author (2025).

electrolyzer modeling methods are summarized in Table 2. Figure 3 shows the power curves of the GE77/1500 and AW125/3000 wind turbines as a function of wind speed. It can be observed that both follow the typical behavior of wind turbines, in which the power increases with wind speed until it reaches the nominal value, remaining constant until the safety cut-off limit. The GE77/1500 model, with a nominal power of 1.5 MW, starts generating energy from approximately 3.5 m/s ( cut-in ), reaching its maximum capacity around 15 m/s ( rated ), and shutting down at 25 m/s ( cut -out). The AW125/3000 wind turbine has a nominal power of 3 MW, with a cut-in at approximately 4 m/s, reaching maximum generation at a lower speed (12 m/s) and shutting down at 25 m/s, which was applied to the state of Ceará. This difference highlights the greater efficiency of the AW125/3000 in utilizing medium-intensity winds, providing nominal power more quickly and at a higher level compared to the GE77/1500. Thus, the joint analysis of the power curves shows that, although both operate within the same safety limit ( cut -out), the AW125/3000 stands out for its greater installed capacity and more favorable performance in intermediate wind regimes, while the GE77/1500 shows a more gradual response to nominal power.

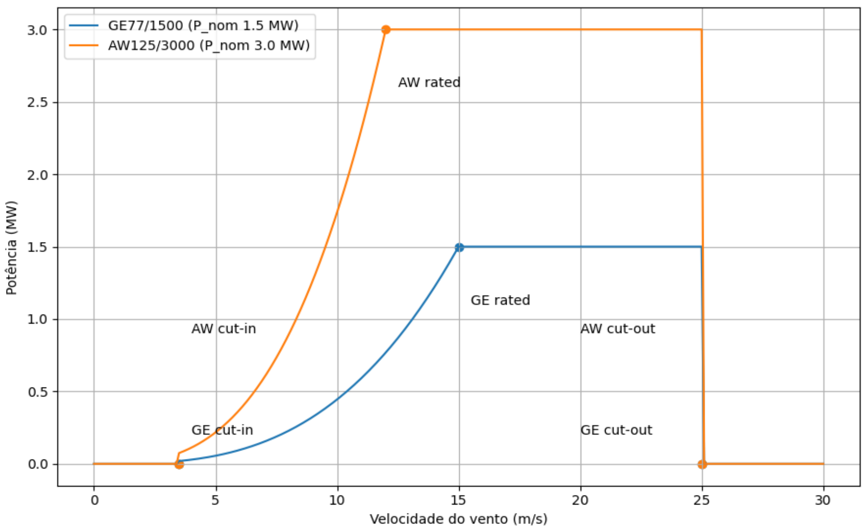



Figure 3: Power curves of the GE77/1500 and AW125/3000 wind turbines

Source: Prepared by the author (2025).

From a technical standpoint, the comparative analysis highlights the importance of choosing the appropriate wind turbine according to the regional wind regime. While the GE77/1500 exhibits a more gradual and robust response up to nominal power, the AW125/3000 stands out for its efficiency in intermediate winds, reinforced by its higher aerodynamic utilization coefficient ( Cp ) and expected superior efficiency in terms of capacity factor described in section 3.1. Table 3 shows the electrolyzer modeling parameters .

**Table** 3 **Similarities and differences between** electrolyzer modeling methods

| **Parameter** | **C1** | **C2** | **C3** | **C4** | **C** |
|---|---|---|---|---|---|
| P_ nom,UEE | 51 MW | 51 MW | 51 MW | 51 MW | 51 MW |
| N_AEG | CE: 17 | CE: 17 | CE: 17 | CE: 17 | CE: 17 |
| η_AEG (%) | CE: 38.84 | CE: 38.84 | CE: 38.84 | CE: 38.84 | CE: 38.84 |
| Power curve | No | Yes | Yes | No | No |
| P_ nom,ELE | 40 MW | 40 MW | 40 MW | 40 MW | 40 MW |
| PCI – H2 | 33.3 kWh/kg | 33.3 kWh/kg | 33.3 kWh/kg | 33.3 kWh/kg | 33.3 kWh/kg |
| η_ELE calibration mode | Numeric | Theoretical | Empirical | Numeric | Numeric |

Source: adapted from Leal, J. et al., (2024).

In all scenarios, the nominal power of the wind farm ( P_nom ,UEE ) is kept constant at 51 MW, as is the nominal power of the electrolyzers ( P_nom,ELE ), fixed at 40 MW, and the lower heating value (LHV) of hydrogen, considered equal to 33.3 kWh/kg, the standard value for gaseous $H_2$ under normal operating conditions. The number of wind turbines (N_AEG) is also identical across scenarios, 17 for Ceará (CE).

The average efficiency of wind turbines ( η_AEG ) in Ceará is 38.84%, reflecting differences in regional anemometric conditions. The main distinction between the scenarios lies in the consideration of the power curve and the method of calibrating the electrolyzer efficiency ( η_ELE ). In scenarios C1, C4, and C5, the power curve is not applied, and the electrolyzer efficiency is adjusted numerically, representing a direct calibration process based on the scenario data. In scenario C2, the power curve is incorporated, and the calibration follows a theoretical method, using analytical relationships between power and hydrogen production. Scenario C3 simultaneously adopts the power curve and an empirical model, allowing efficiency to be estimated from experimental or historical data.

# 4. RESULTS and DISCUSSION

## 4.1 Results of the Annual and Daily Hydrogen Production Forecast for the Scenarios of the State of Ceará

The results obtained for Ceará reveal significant differences between scenarios C1 to C5, and these differences are directly explained by the conversion methods adopted. In Scenario 1, an average production of 6.63 million kg/year (≈ 18,177 kg/day, ≈ 757 kg/h) is observed, with an average power of 25.22 MW and a capacity factor of 0.63. This high performance stems precisely from the stationary approach, which assumes idealized operation of the electrolyzer .

Scenario 2 reduces the average annual production to 3.49 million kg/year (≈ 9,561 kg/day, ≈ 398 kg/h). This lower value is consistent, as the method considers real losses in the load ranges, especially at low loads, which reduces the useful power delivered to the electrolyzer . The average power drops to 13.27 MW, with FC = 0.33.

In Scenario 3, production of 4.84 million kg/year (≈ 13,253 kg/day, ≈ 552 kg/h), with 18.39 MW and FC = 0.46. The continuous method avoids the abrupt yield jumps present in C2, generating a smoother result according to the efficiency of the electrolyzer .

Scenario 4 presents 4.03 million kg/year (≈ 11,038 kg/day, ≈ 460 kg/h), with 15.32 MW and FC = 0.38. As the method considers aerodynamic losses, clipping regions and real blade behavior, the value is slightly lower than the more idealized scenarios such as C1 and C5.

Finally, Scenario 5 presents productions of 8.20 million kg/year (≈ 22,456 kg/day, ≈ 936 kg/h), with 31.16 MW and FC = 0.78. Considering that the temperature and pressure for calculating air density used in this model are not real data, but idealized averages, as in other compared states, a difference in the volume produced is noted compared to scenario 1, where constant air density was considered.

Another point to consider is the daily production figures presented, as the values may differ for each day considered, since wind speeds vary, as do production figures along with the range limits considered in the C2 and C3 models, which will be discussed further in this work.

These results demonstrate the importance of incorporating realistic physical parameters and load-dependent efficiency functions in the electrolysis process modeling, especially when the goal is to predict green hydrogen production on a regional scale. The increasing and decreasing trend observed between scenarios indicates that, as the modeling approaches real operating conditions, the projections become more robust and consistent with wind power potential.

Figure 4 shows the production forecasts; the series demonstrate a strong temporal correlation, reflecting the direct influence of wind variability on the electrolysis rate. In general, a high regularity of wind power generation is observed throughout the year , with small seasonal fluctuations. The months between June and September concentrated the highest generation levels, corresponding to the season of most intense winds on the Ceará coast, while December showed a slight reduction. This constancy of the wind regime guarantees a continuous and predictable energy supply for the electrolysis system.

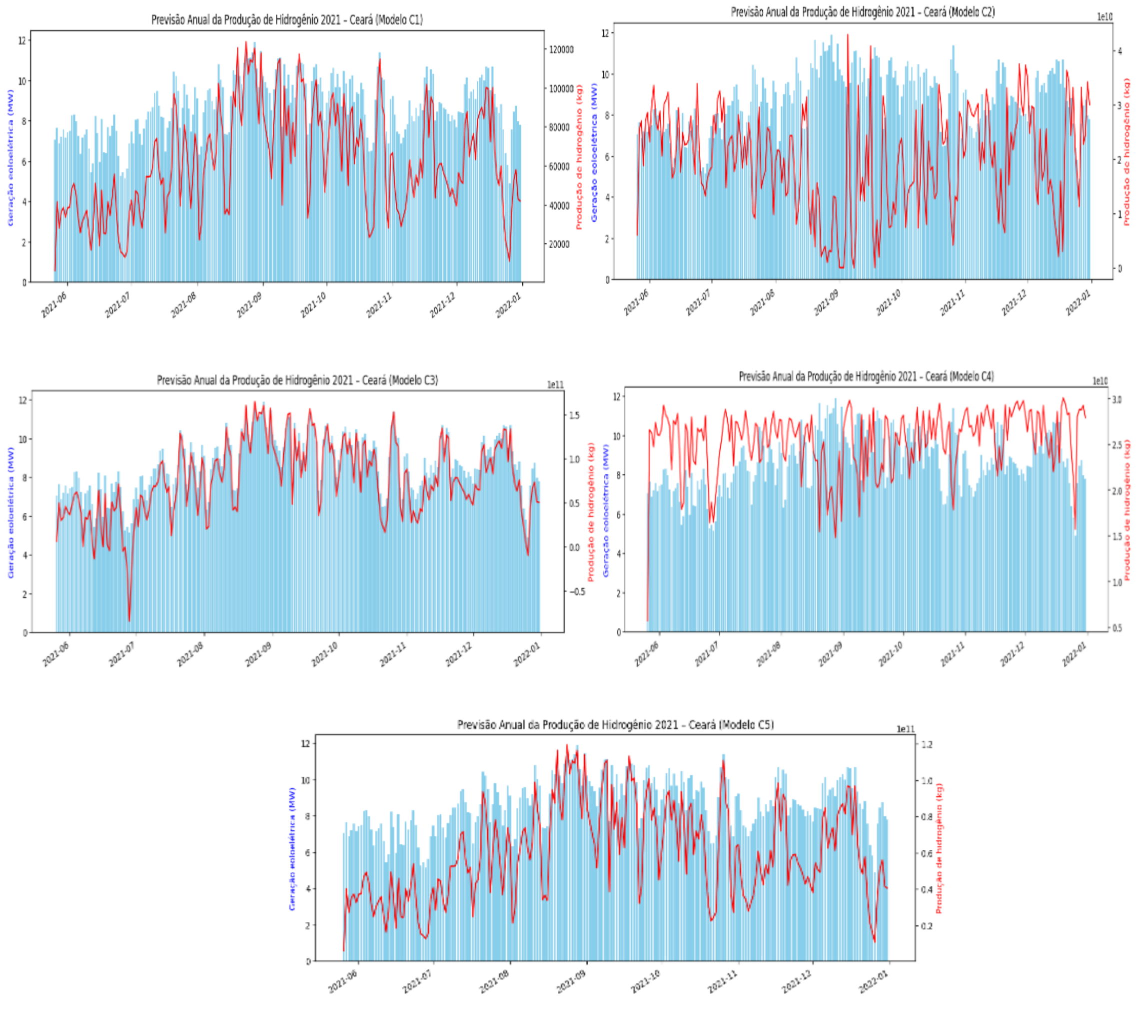



**Figure 4 –** Annual forecast of hydrogen production in Ceará, 2021

Source: Prepared by the author (2025).

In scenarios C2 and C3, it is observed that the production ranges have well-defined limits, especially at the lower and upper ends. This is because the operational limits of the electrolyzers were explicitly applied in these scenarios , i.e., the minimum and maximum power ranges allowed for the electrolysis process. In these scenarios, the system operates within predefined minimum and maximum load ranges (e.g., 40–100% for ALK or 10–125% for PEM), which prevents hydrogen production below the lower power limit and avoids indefinite growth above the upper limit. As a consequence, hydrogen production never exceeds the rate corresponding to the nominal power nor falls below the minimum operating threshold, creating a rigid physical range for the predicted values. Unlike scenario C4, which uses the efficiency curve of the wind turbines, these limits do not appear abruptly. Instead, the production behavior results from the shape of the turbine power curve itself, which introduces a smooth transition between the low-efficiency regions, the cubic region, and the constant power region. Thus, the limits observed in C4 directly reflect the physics of wind energy conversion, and not an explicit imposition of restrictions, producing more continuous forecast ranges that are more compatible with the aerodynamic profile of wind turbines.

The energy assessment of the wind→$H_2$ conversion was carried out by calculating the specific electrical consumption and the average electrical power required for each scenario, adopting the lower calorific value of hydrogen (LHV) = 33.3 kWh/ kg , and considering the nominal efficiencies of the electrolyzer defined in the scenarios. It was found that the actual electrical consumption per kilogram of $H_2$ varies between approximately 40.6 kWh/kg (favorable partial load condition, 82%) and approximately 50.5 kWh/kg (overload condition, 66%), while for electrolyzers operating at 75% efficiency, consumption tends to **approximately** 44.4 kWh/kg . From the estimated average daily production, the daily electrical energy required was calculated and, subsequently, the average continuous electrical power (MW) needed to sustain production. For example, in Ceará (C1) the average production of 18,177 kg/day ≈ 757 kg/h implies an average electrical input of approximately 33.6 MW (equivalent to ≈807 MWh/day), resulting in an average $H_2$ output power (LHV) of 25.2 MW . In terms of chain efficiency ( wind→$H_2$ ), and without explicitly recalculating the aerodynamic efficiency of the turbine, the overall efficiency can be estimated as $\eta tot \approx$ 25% . This order of magnitude is consistent with recent literature and shows that a significant fraction of the raw wind resource is lost in the conversion and balance- of - plant stages . Consequently, for desired production levels in the order of 18-23 t/day , the associated electrical infrastructure must provide between ~ 33-43 MW of average electrical power (depending on the actual efficiency of the electrolyzer ), which reinforces the need for grid integration strategies, storage, or oversizing of the wind farm to ensure operational continuity under resource variability.

### 4.1.2 Results of the Metrics for Evaluating CE Scenarios

For the test, C1 RMSE of 201.95 kg/km $^2$ and MAE of 155.12 kg/km² in the set, with a MAPE of 29.72%. The $R^2$ of 0.7205 indicates that approximately 72% of the data variability was explained by the model. The Skill Score was 0.7094 and the training time was 990 seconds. C2, with an RMSE of 86.33 kg/km² , MAE of 69.39 kg/km² , MAPE of 24.02%, an $R^2$ of 0.5431, a Skill Score of 0.7683, and a training time of 649 seconds. C3 Test $R^2$ 0.7293 and a Skill Score of 0.6879, with an RMSE of 280.53 kg/km² and a MAE of 216.44 kg/km² , a MAPE of 35.53%, and a training time of 865 seconds. C4, RMSE 45.19 kg/km² , MAE 31.72 kg/km² and Skill Scores 0.8316, $R^2$ 0.4928 and training time 692 seconds. C5 With an RMSE of 220.74 kg/km² , MAE of 158.95 kg/km² , and MAPE of 30.52%, and an $R^2$ of 0.7191, the Skill Score was 0.6797, with a training time of 1131 seconds, as shown in Table 5.

Table 5: The main metrics for comparing errors between scenarios.

| **Metric** | **C1** | **C2** | **C3** | **C4** | **C** |
|---|---|---|---|---|---|
| **Best number of lags** | 10 | 8 | 14 | 5 | 14 |
| **Training time (s)** | 990.24 | 649.02 | 865.64 | 692.44 | 1130.76 |
| **Training** | | | | | |
| RMSE ( kg/ km2 ) | 165.37 | 81.95 | 240.44 | 46.72 | 167.14 |
| MAE (kg/km$^2$) | 116,77 | 63,03 | 168,81 | 34,82 | 108,65 |
| MAPE (%) | 38,99 | 32,99 | 40,91 | 27,77 | 51,84 |
| R² | 0,8035 | 0,7010 | 0,7857 | 0,5836 | 0,8237 |
| Skill Score | 0,6449 | 0,7016 | 0,6067 | 0,8114 | 0,5994 |
| **Teste** | | | | | |
| RMSE (kg/km$^2$) | 201,95 | 86,33 | 280,53 | 45,19 | 220,74 |
| MAE (kg/km$^2$) | 155,12 | 69,39 | 216,44 | 31,72 | 158,95 |
| MAPE (%) | 29,72 | 24,02 | 35,53 | 30,36 | 30,52 |
| R² | 0,7205 | 0,5431 | 0,7293 | 0,4928 | 0,7191 |
| Skill Score | 0,7094 | 0,7683 | 0,6879 | 0,8316 | 0,6797 |

Source: Prepared by the Author (2025).

While Model 4 stands out for its physical fidelity and absolute accuracy, Model 2 shows advantages when seeking a balance between performance, processing time, and adaptability. Both models demonstrated distinct advantages, making them more suitable depending on the application's ultimate goal. Figure 5 provides a better visualization comparing the scenarios.

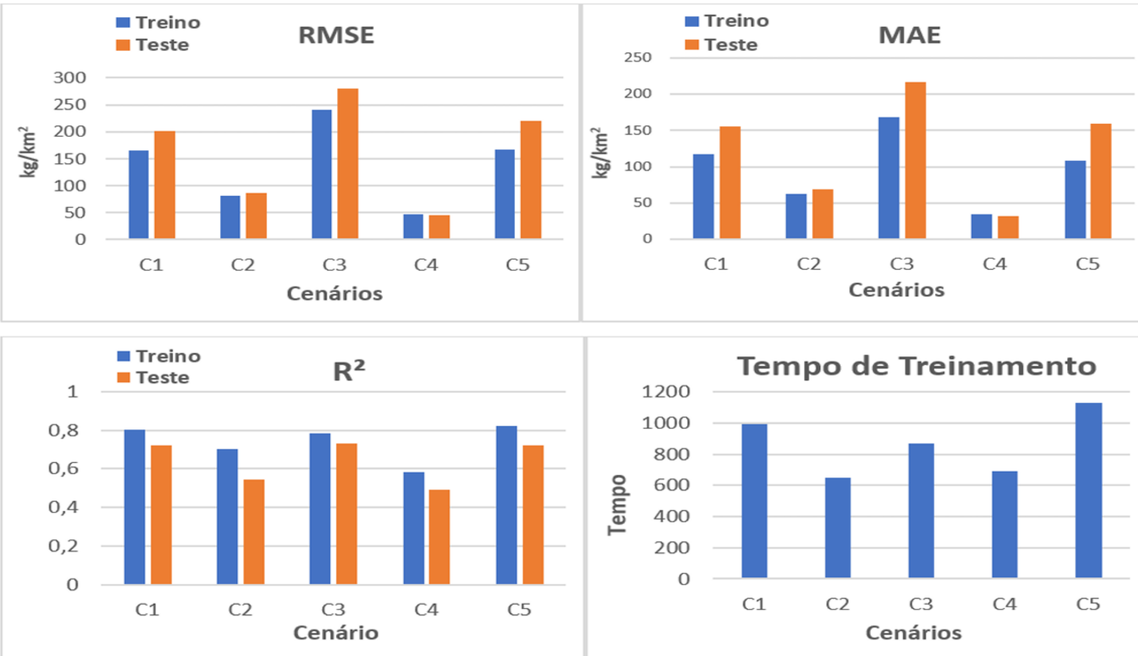



Figure 5: Comparison of the scenarios.

Source: Prepared by the Author (2025).

C3, the sophisticated polynomial modeling, likely introduced noise and caused instabilities, making it difficult to generalize its learnings to new data.

## Acknowledgments

This study was financed in part by FUNCAP - Ceará Foundation for Support of Scientific and Technological Development ( Research) and Innovation Network on Renewable Energy, Rede VERDES, Grant No. 07548003/2023), CNPq - National Research Council, Universal - Grant No. 408431/2023-7 and CNPq - National Research Council, Research Productivity Grant No. 303585/2022-6 and 307096/2021-1.

## 4 FINAL CONSIDERATIONS

The work developed offers a robust framework for estimating green hydrogen production based on real meteorological data, combining physical modeling and artificial intelligence. Five scenarios were addressed, each incorporating different hypotheses about the efficiency and behavior of the wind-to-hydrogen conversion system. Thus, it is concluded that the use of machine learning models, combined with realistic representations of physical processes, is a promising approach for predicting green hydrogen production, offering technical and scientific support.

## 5 REFERENCES

ABASIMEHR, H.; PAKI, R.; BAHRINI, A. A novel XGBoost-based characterization approach for predicting renewable energy consumption with deep learning models. *Sustainable Computing: Informatics and Systems*, v. 38, n. 4, p. 1–14, 2023.

ALLAL, Z.; NOURA, H. N.; SALMAN, O.; CHAHINE, K. Machine learning solutions for renewable energy systems: applications, challenges, limitations and future directions. *Journal of Environmental Management*, 354:120392, 2024. https://doi.org/10.1016/j.jenvman.2024.120392

BRAZIL WINDPOWER PAPERS. Green industrial policy and just energy transition: the Brazilian leadership role. BWP Digital Platform, 2023.

CARMO, M.; FRITZ, DL; MERGE, J.; STOLTEN, D. A comprehensive review on PEM water electrolysis. *International Journal of Hydrogen Energy* , vol. 38, no. 12, p. 4901–4934, 2013.

FARKAS, Z. Considering air density in wind energy production. Department of Physics of Complex Systems, Eötvös University , Budapest, Hungary, 2011.

GUITARRARA, P. “Hidrogênio verde”. *Brasil Escola*. Disponível em: https://brasilescola.uol.com.br/geografia/hidrogenio-verde.htm. Acesso em: 19 jun. 2025.

HOFRICHTER, A. et al. Determination of the optimal power ratio between electrolysis and renewable energy to investigate the effects on the hydrogen production costs. *International Journal of Hydrogen Energy*, v. 48, n. 5, p. 1651–1663, 2023.

JAVAID, A. et al. Forecasting hydrogen production from wind energy in a suburban environment using machine learning. *Energies*, v. 15, n. 23, 2022.

KUMAR, V.; PAL, Y.; TRIPATHI, M. M. PSO-tuned ANN-based prediction technique for penetration of wind power in grid. In: *Proceedings of the International Conference on Recent Innovations in Computing (ICRIC)*. Springer, 2019. p. 485–494.

LIU, Y. et al. Degradation of nickel electrodes under power fluctuations in alkaline electrolyzers. *Journal of Materials Chemistry A*, v. 12, n. 3, p. 1432–1444, 2024.

MANWELL, J. F.; McGOWAN, J. G.; ROGERS, A. L. *Wind Energy Explained: Theory, Design and Application*. 2nd ed. John Wiley & Sons, 2009.

THOMAS, J. et al. Wind Turbine Design Optimization for Hydrogen Production. National Renewable Energy Laboratory (NREL), Ponta Delgada, Açores, Portugal, 2024.

ZHU, L. et al. Optimal operation strategy for power-to-hydrogen system considering dynamic efficiency. In: *2022 IEEE 6th Conference on Energy Internet and Energy System Integration (EI2)*. Chengdu, 2022. p. 836–842.