



HAL
open science

Data-driven approach and machine learning for intelligent control of alkaline electrolyzers

Ismail Ramde, Jesus Vital Kombaya Touckia, Sébastien Henry, Yacine Ouzrout,
Hector Rivera

► **To cite this version:**

Ismail Ramde, Jesus Vital Kombaya Touckia, Sébastien Henry, Yacine Ouzrout, Hector Rivera. Data-driven approach and machine learning for intelligent control of alkaline electrolyzers. The 52nd International Conference on Computers and Industrial Engineering (CIE52), Oct 2025, Lyon, France. <hal-05597037>

HAL Id: hal-05597037

<https://hal.science/hal-05597037v1>

Submitted on 20 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

Data-driven approach and machine learning for intelligent control of alkaline electrolyzers

Ismail RAMDE^{a,b}, Jesus KOMBAYA TOUCKIA^b, Sébastien HENRY^b, Yacine OUZROUT^b, Hector RIVERA TREVIÑO^a

^aMcPhy Energy, Grenoble, France

^bINSA Lyon, Université Lumière Lyon 2, Université Claude Bernard Lyon 1, Université Jean Monnet Saint-Etienne, DISP UR4570, Villeurbanne, 69621, France

Abstract

In the context of climate change, the transition to clean energy sources such as green hydrogen is essential to reduce dependence on fossil fuels. This study was conducted on a one megawatt industrial alkaline electrolyzer that has been in operation for over two years. Given the complexity of the system and the lack of studies based on real data, we propose an approach based on machine learning and integrated into a robust modeling pipeline that can anticipate liquid level deviations in the separators. This imbalance causes excessive electricity consumption and disrupts pressure and flow. Several models were trained using a large amount of data. The XGBoost model, with a coefficient of determination equal to 0.99, proved to be the best compromise. It will be integrated into the control logic to enable the liquid level control valve to react in advance. This intelligent control will improve energy efficiency.

Keywords: Machine learning, Separation level, Energy optimization, Smart regulation, Hydrogen production, Alkaline electrolyzer

1. Introduction

Over the past decade, research into energies that could promote ecological transition, particularly hydrogen, has accelerated. This growth is supported by government efforts, the formation of multiple international coalitions, and the implementation of several projects. Hydrogen is now emerging as a promising solution thanks to its massive storage potential and zero carbon dioxide (CO_2) emissions. Although the chemical process is simple - splitting a water molecule - an electrolyzer is a complex system. Controlling the process over long periods of operation in the face of chemical kinetics, as well as managing pressure, temperature, and flow rates, are major challenges. Voltage management and stack component degradation also disrupt the expected behavior of the system.

Energy efficiency, operational stability, and the ability to adapt to load variations are key performance indicators that need to be optimized. For a fleet of electrolyzers, it is essential to improve these performances by acting on the control system. Conventional methods such as PID (Proportional, Integral, Derivative) controllers are not capable of anticipating sudden system variations. This can be achieved using predictive approaches based on artificial intelligence (AI) that can anticipate undesirable phenomena. There is a growing interest in AI methods to optimize the energy performance of electrolyzers [1, 2].

The work carried out in the article [3] focused on predicting production potential based on climate data, solar radiation, and temperature. Research based on CFD (Computational Fluid Dynamics) simulations and neural networks has shown that geometric design and ambient conditions can strongly influence the performance of proton exchange membrane (PEM) electrolyzers [4]. One study involved the integration of a photovoltaic thermal system with a PEM electrolyzer. The model employs CatBoost to predict and optimize hydrogen production [5]. An approach

Email addresses: ismail.ramde@etu.univ-lyon1.fr (Ismail RAMDE), jesus.kombaya-touckia@univ-lyon1.fr (Jesus KOMBAYA TOUCKIA), sebastien.henry@univ-lyon1.fr (Sébastien HENRY), yacine.ouzrout@univ-lyon2.fr (Yacine OUZROUT), hector.rivera@mcphy.com (Hector RIVERA TREVIÑO)

combining machine learning and thermal balance allows the operating parameters of alkaline electrolyzers to be adjusted, reducing energy costs by 10 [6]. The use of CFD simulation and neural networks to model the performance of an alkaline electrolyzer, taking into account parameters such as temperature, electrolyte concentration, and cell geometry, has enabled accurate prediction of the impact of these parameters on efficiency [7]. Improved energy efficiency and a 20% reduction in production costs were achieved by exploring the optimal operating conditions of electrolyzers using neural networks and reinforcement learning for green hydrogen production [8]. The authors in the article [9] proposed a combination of multi-objective optimization, XGBoost, and genetic algorithms to improve hydrogen production from methane reforming. This reduced costs and CO_2 emissions while increasing yield. An energy optimization model was developed to improve hydrogen production through methane reforming. Designed based on a neural network, this model achieved a thermal efficiency of 85.5% [10]. The study cited in the article [11] explored the use of machine learning to optimize PEM electrolyzers for green hydrogen production. The Cascade Forward Neural Network (CFNN) model proved to be the most effective for predicting cell voltage. A proposal for Support Vector Machine (SVM) and Fbprophet models, based on climate data, made it possible to predict green hydrogen production from PEM and alkaline electrolyzers powered by a solar energy source. Good performance was recorded for areas with high solar radiation [12]. Finally, the authors of the study [13] used predictive models such as XGBoost to optimize the performance of Solid Oxide Electrolytic Cell (SOEC) electrolyzers. They were able to identify the optimal conditions.

These studies demonstrate the growing role of AI in the energy optimization of electrolyzers. Furthermore, all these approaches are based on data generated by simulation from theoretical models of electrolyzers, with the exception of the study [14], which is based on a 4.5 kW laboratory electrolyzer. The theoretical approaches are based on ideal assumptions and do not reflect the dynamic effects (pressure and flow peaks observed in the real system) of the electrolyzer. The predictive models derived from these approaches are less reliable than those of the industrial electrolyzer in a real production environment. In most cases, they aim to find optimal static operating parameters rather than adaptive control through piloting.

In this article, we used real data from a 1 megawatt (MW) alkaline electrolyzer that has been in production for over two years to develop and integrate a predictive model into the electrolyzer's control loop.

Our study consists of three parts. The first part focuses on the methodology. The second part presents the results of the case study and analyzes them. Finally, the last part presents the conclusions of the study.

2. Methodology

2.1. Description of an alkaline electrolyzer

For confidentiality reasons, illustrations and descriptions of the system will be limited to general information. Sensitive data will also be concealed.

The alkaline electrolyzer produces hydrogen by alkaline water electrolysis (AWE) at temperatures typically between 60°C and 90°C [15]. It generally uses a potassium hydroxide (KOH) solution. Unlike other technologies, AWE is the most mature. It has the advantage of a longer service life, low manufacturing costs, and higher hydrogen purity [16].

Fig. 1 illustrates the main components of an alkaline electrolyzer. The stack is the central element where chemical reactions take place. It consists of the anode, the cathode, and a porous membrane that separates hydrogen (H_2) and oxygen (O_2). A pumping system injects the aqueous KOH solution (electrolyte) into the stack. The electric current injected into the stack activates the dissociation of the water molecule. A network of pipes, valves, sensors, and auxiliary equipment, as well as pressurized storage, complete the system.

2.2. System Modeling and Framework

System modeling began with an analysis of the Piping and Instrumentation Diagram, which proved too complex for effective data exploitation and difficult to interpret. The challenge for us was to move from a complex representation of the electrolyzer and a large amount of raw data to a detailed and actionable understanding of the electrolyzer in order to optimize its energy efficiency. We therefore simplified this diagram to produce a data-driven model. We then adopted a modular approach by dividing the system into subsystems with their inputs and outputs. This division

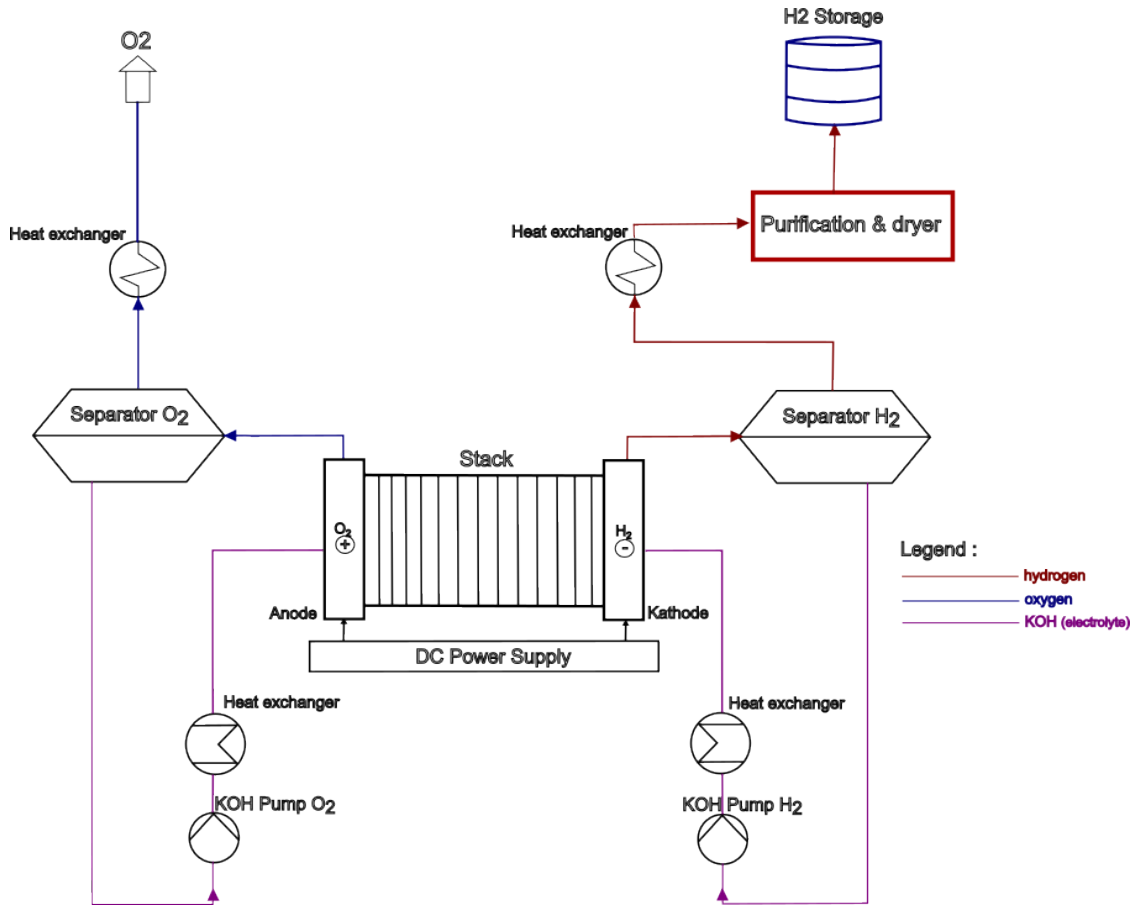


Figure 1: Alkaline Water Electrolyzer

allows for a better understanding of the interactions between the subsystems and an interpretation of the causal links. In addition, the modular approach allows us to propose both global and local optimizations for each subsystem.

Fig. 2 describes our optimization framework, designed according to the modular approach. It includes structuring, data processing, and machine learning. The process also integrates exploratory analysis, model creation, and evaluation. The models are put into production after the testing and validation phases.

2.3. Data preparation and processing

Electrolyzers generate a large volume of data. This data consists of 169 variables logged at a frequency of one measurement per second. Thus, for data ranging from May 2023 to December 11, 2024, a total of 5,024,229 lines of data were collected. A selection of 66 relevant variables was made using internal engineering knowledge, focusing on production cycles. The cycles were constructed based on the status of a sensor indicating the production status (0 = stopped, 1 = running).

Data processing involved several steps: cleaning, applying filters, correcting outliers and missing values using the linear interpolation method. A feature engineering phase was used to create derived variables and calculate key performance indicators:

- the difference in liquid and pressure levels:

$$|\Delta_{Level}| = |Level_{H_2} - Level_{O_2}| \quad (1)$$

where:

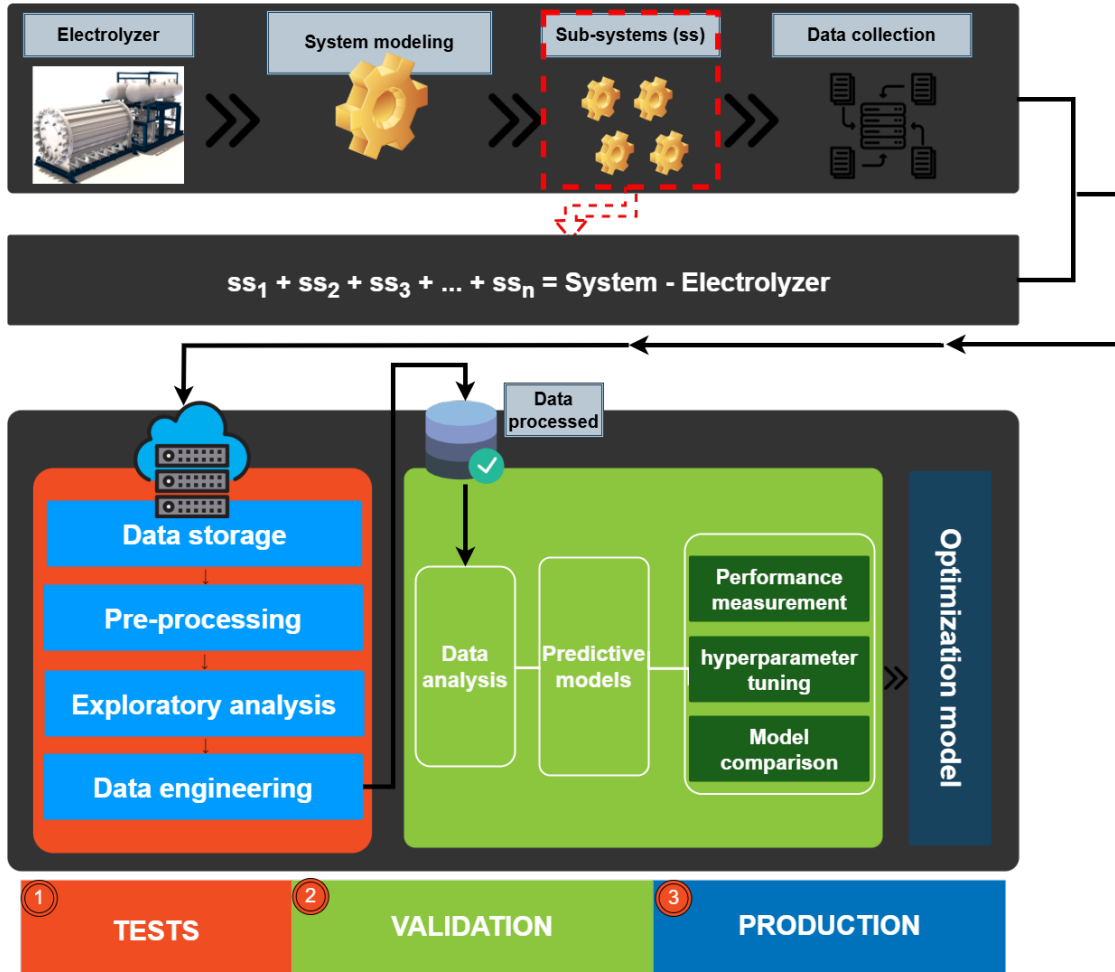


Figure 2: Predictive modeling framework

Δ_{Level} is the difference between the liquid levels, $Level_{H_2}$ is the measured level of the H_2 separator (in %) and $Level_{O_2}$ is the measured level of the O_2 separator (in %).

- Power :

$$P_{stack} = I_{total} \times V \quad (2)$$

where:

P_{stack} is the electrical power of the stack (in W) which will be converted into kWh, I_{total} is the electrical current applied in amperes (A) and V is the voltage in volts (V).

- Energy efficiency :

$$Energy\ Efficiency = \frac{Power\ Consumed}{kg\ H_2\ Produced} \quad (3)$$

where:

Power Consumed is the energy consumed (in kWh), kg H_2 Produced is the mass of hydrogen produced (in kg).

When Faraday's efficiency (η_F) is taken into account, the formula becomes:

$$Energy\ Efficiency_{Faraday} = \frac{Power\ Consumed}{kg\ H_2\ Produced \times \eta_F} \quad (4)$$

with η_F : Faraday efficiency (usually expressed as a percentage or decimal).

2.4. Data analysis

Careful analysis of the data obtained from this processing stage enabled us to extract important information relating to the operation and performance of the electrolyzer. The modular approach we adopted (see Fig. 2) enabled us to identify an area for improvement in the management of the liquid level (mixture of gas and liquid) contained in the H_2 and O_2 separators. It is essential to maintain a balance between these two levels to ensure optimal operation of the electrolyzer. To this end, the system is equipped with a control valve capable of opening gradually to restore the balance. It is activated and controlled by a PID controller, which struggles to anticipate significant variations and adapt to dynamic conditions due to its type of control.

We were then able to evaluate the impact of this valve control on the performance of the electrolysis using measurement tools, including Pearson (r) and Spearman (ρ) correlations. These tools measure the degree and direction of the relationship between two variables.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

where: r is Pearson's correlation coefficient, x_i, y_i are the observed values of the variables, \bar{x}, \bar{y} are the means of x and y , n is the number of observations.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (6)$$

where:

(ρ) is the Spearman correlation, d_i is the difference between the ranks of the observations of x_i and y_i , n is the number of observations.

2.5. Predictive modeling using machine learning

In this section, we present the different machine learning models that have been implemented and evaluated to predict the difference in liquid level (Δ_{Level}) in the H_2 and O_2 separators. The predictions of the best model will then be used to feed into the intelligent valve control strategy to maintain the liquid balance.

2.5.1. Selection of relevant explanatory variables

The variable we are trying to predict at a 10-second time horizon is the difference in liquid levels Δ_{Level}_{t+10} . The explanatory variables were selected in two stages. First, we correlated all the variables with the variable Δ_{Level} to produce a list of variables. Then, we combined our internal engineering knowledge to select the most relevant ones. The list of variables selected is as follows:

- current (x_1)
- production load (x_2)
- power (x_3)
- voltage (x_4)
- temperature (x_5)
- Δ_{Level} , its moving average and standard deviation over 30 seconds (x_6 and x_7, x_8)

2.5.2. Machine learning models

The prediction of fluid level imbalance $\Delta Level_{t+10}$ was performed using several machine learning models. We chose to use models from different families in order to have a wide range of options. Four main model families were used, as shown in Table 1, which compares the different models and summarizes the essential information for each one.

A linear model was used as the baseline model. Non-linear models, such as Random Forest, XGBoost, and CatBoost, were developed for their ability to capture non-linear and complex relationships. Neural networks, whether artificial (ANN) or deep (DNN), are known for their ability to handle instabilities and non-linear relationships. The time series model, Prophet, was also tested. Time series are well suited to modeling the temporal dynamics of the system.

Table 1: Comparison of models applied to the prediction of $\Delta Level_{t+10}$

Criterion	Linear Regression	Random Forest	XGBoost	CatBoost	ANN	DNN	Prophet (Time Series)
Model type	Linear	Ensemble of trees	Tree boosting	Boosting with cat.	Neural network	Deep neural network	Additive model (time)
Complexity	Low	Medium	High	High	High	Very high	Medium
Model structure	Fine-grained	Average tree	Serial tree addition	Boosting + cat. encoding	Dense layers	Deep layers	Trend, seasonality, holidays
Non-linearity handling	No	Yes	Yes	Yes	Yes	Yes	Partially
Scalability	Good	Good	Excellent	Excellent	Excellent	Very good	Good
Interpretability	Very good	Average	Average	Average	Poor	Poor	Good
Learning speed	Very fast	Average	Slow	Slow	Slow	Slow	Average
Test speed	Very fast	Fast	Fast	Fast	Fast	Fast	Average
Strengths	Simplicity, interpretable	Robust, few adjustments	High performance, regularized	Very good performance	Powerful, adaptable	Very flexible	Good for time series
Limitations	Too simple	Less interpretable	Complex tuning	Sensitive to overfitting	Requires tuning	Risk of overfitting	Less suitable for noisy signals
Model formula	$f(X) = \beta_0 + \sum \beta_i X_i$	$f(X) = \frac{1}{T} \sum h_t(X)$	$f(X) = \sum \eta h_t(X)$	Same as XGBoost + cat. encoding	$f(X) = W_2 \sigma(W_1 X + b_1) + b_2$	$f(X) = W_n \cdots \sigma(W_1 X + b_1)$	$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$

where:

- x_i : i^{th} explanatory variable,
- β_0, β_i : bias term and coefficients in the linear model,
- T : total number of trees (Random Forest, Gradient Boosting),
- $h_t(X)$: prediction of the t^{th} tree,
- η : learning rate (Gradient Boosting),
- W_i, b_i : weights and biases of the layers in the neural network,
- σ : activation function (ReLU, tanh, etc.),
- $y(t) = g(t) + s(t) + h(t) + \varepsilon_t$: components of the Prophet model with $g(t)$ the trend, $s(t)$ the seasonality, $h(t)$ the events, and ε_t the random noise.

2.5.3. Criteria for comparing models

To choose the best model, we used two performance indicators: the coefficient of determination (R^2) and the Root Mean Squared Error (RMSE). These indicators measure the accuracy of the models. The coefficient of determination measures the quality of the fit. The RMSE measures the standard deviation of the prediction errors.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

where:

y_i is the actual value of the variable to be predicted for observation i , \hat{y}_i is the value predicted by the model for observation i , \bar{y} is the average of the actual values y_i and n is the total number of observations.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

where: the symbols y_i , \hat{y}_i , and n are defined as in the previous equation.

3. Case study and discussion

This research is being conducted in an industrial environment. It focuses on alkaline electrolyzers with a unit power of 1 MW under real production conditions. These electrolyzers are installed in several cities in France.

3.1. Pressure analysis

To evaluate the effect of pressure rise on the energy efficiency of the electrolyzer, our results are compared with those presented by Koponen et al [14]. Their study is based on a laboratory PEM electrolyzer.

During production, the pressure rises gradually to reach 30 bar. The electrolyzer spends very little time in this pressure rise phase. Thus, more than 98% of the measurements are concentrated at pressures above 20 bar. In fact, pressure ranges below 20 bar are not statistically significant in relation to the whole.

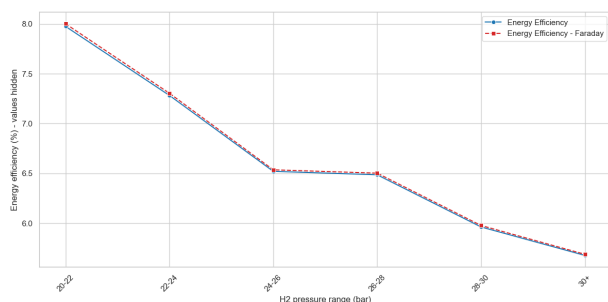


Figure 3: Energy efficiency according to H2 outlet pressure

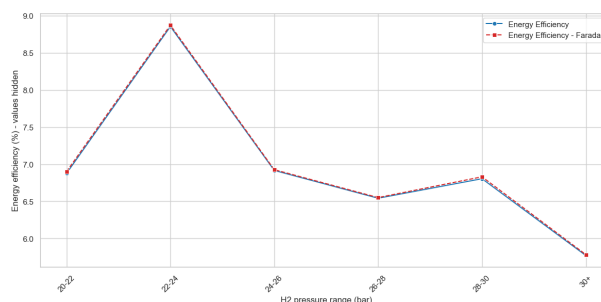


Figure 4: Energy efficiency according to H2 pressure during the process

To analyze the effect of pressure, we divided it into intervals of 2, ranging from 20 to 30 bar. We then plotted the average curve for each interval as a function of the electrolyzer's energy efficiency, as shown in Figs. 3 and 4. The first represents the hydrogen outlet pressure and the second represents the hydrogen pressure during the molecule dissociation process.

The two curves shown in each figure (energy efficiency and energy efficiency taking into account Faraday efficiency) show a downward trend. This means that the electrolyzer is efficient at high pressures, ideally 30 bar. This result differs from that obtained in the previous study, which showed relative stability in terms of pressure. On the other hand, a very slight increase in energy efficiency taking into account Faraday efficiency (in red) is noticeable.

3.2. Pressures and liquid levels in separators

The separation of H_2 and O_2 takes place within separators designed for this purpose. For optimal operation, it is essential that the difference in liquid level between the two separators be as close to zero as possible. This regulation is ensured by a valve with gradual opening controlled by a PID regulator.

The separations also generate pressure in each compartment, on the H_2 side (seen above in Fig. 4) and O_2 , which must also remain in equilibrium.

Our analyses show that the PID controller in place is unable to anticipate level fluctuations, resulting in significant peaks in valve behavior, as illustrated by the orange curve in Fig. 5 of a production cycle, which reflects the same phenomenon across all cycles. In addition, a moderate correlation of -0.19 (Spearman) was observed between the liquid level deviation and the pressure deviation of the separators. The system is unbalanced due to the significant deviations observed, which has an effect on power consumption. Indeed, the correlation with pressure imbalance is +0.44, while that with liquid level imbalance is -0.22. These results show that good balance management can improve the efficiency of the electrolyzer.

3.3. Prediction of level differences

3.3.1. H_2 and O_2 level control valve

Our smart H_2 and O_2 level control strategy involves anticipating level deviations in order to maintain balance. We have therefore implemented and tested machine learning models that enable us to accurately predict imbalances 10 seconds in advance.

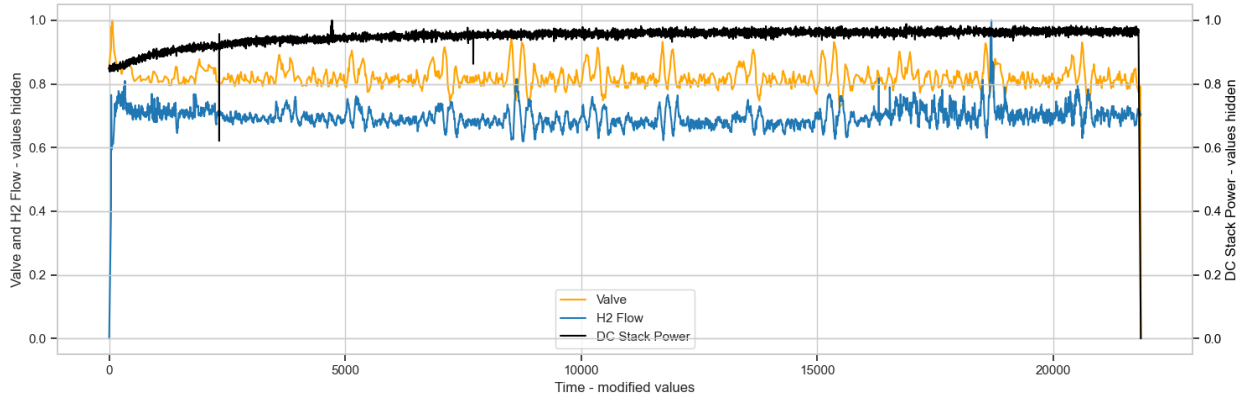


Figure 5: Production = 210: illustration of variations in the valve, H_2 Flow and Power

These predictions will feed into the PID controller, which will now be able to anticipate regulation by opening very slightly. This will significantly reduce sudden openings and peak amplitudes.

3.3.2. Comparison of model performance

Table 2 presents the performance results of the different models we developed to predict $\Delta Level_{t+10}$. The models are evaluated based on R^2 and RMSE. We can see that all models perform very well, with values between 0.86 and 0.99. This reflects a very good ability to explain the variance of the variable to be predicted. However, there are significant differences in the RMSE values.

Table 2: Model performance

Model	R^2	RMSE
Linear regression	0.98	0.40
Random Forest	0.99	0.10
XGBoost	0.99	0.10
CatBoost	0.99	0.12
DNN	0.99	0.14
ANN	0.99	0.14
Prophet	0.86	1.04

The linear regression model has an R^2 score of 0.98 and an RMSE of 0.40. It is easy to interpret and fast compared to other models, but remains limited in modeling nonlinear relationships.

The Random Forest, XGBoost, and CatBoost models, based on trees, perform better with $R^2 = 0.99$ and RMSEs between 0.10 and 0.12. The Random Forest and XGBoost models are the most accurate with RMSEs of 0.10 and are robust to variations in the data.

The Prophet model shows less interesting production capacity with an $R^2 = 0.86$ and an RMSE = 1.04.

As for neural networks (DNN and ANN), they have $R^2 = 0.99$ and RMSE = 0.14. This demonstrates excellent prediction quality similar to that of tree-based models.

The models we implemented perform very well, with predictions very close to the actual data. The XGBoost model has very high accuracy ($R^2 = 0.99$, RMSE = 0.10). Figure 6 illustrates the plot of the actual data and the predictions of the XGBoost model on a sample of 300. The graph shows that the difference between the two curves is very small. The model is able to capture the dynamics of the system very well.

It stands out for its robustness to noise and extreme values. In our industrial context of real-time regulation, its reduced inference time is a major advantage. It also offers good management of non-linearity, unlike the linear model. Despite these advantages, its execution time is slightly longer than that of the Random Forest model and it is less interpretable than the linear model. This is why we use the SHAP (SHapley Additive exPlanations) method,

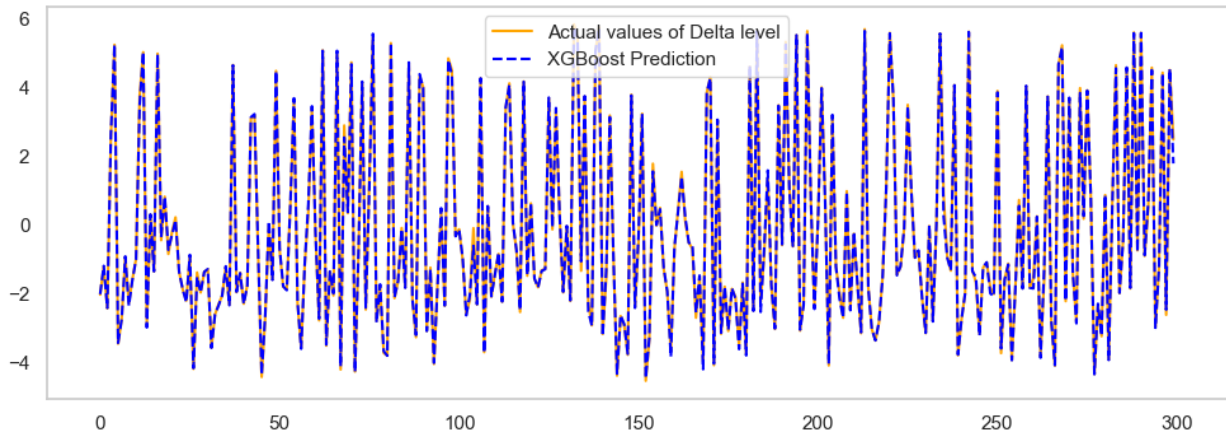


Figure 6: XGBoost - Prediction values vs. Actual values

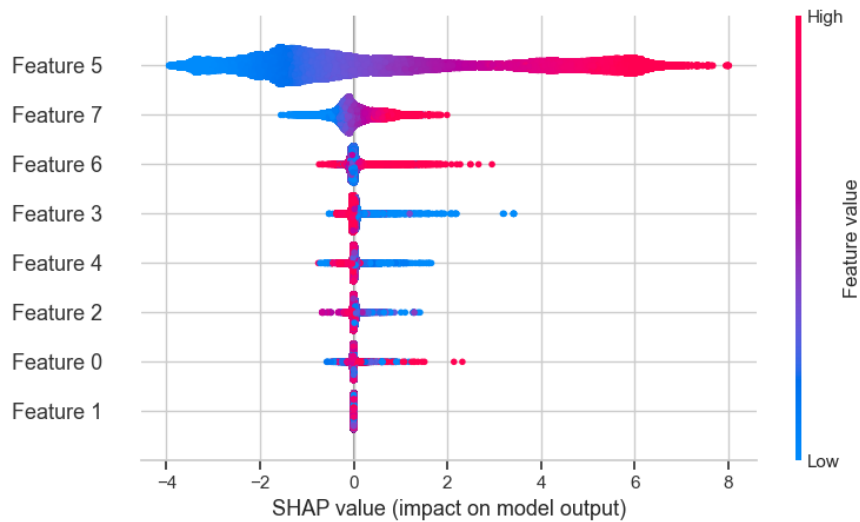


Figure 7: Interpretation of explanatory variables using SHAP

illustrated in Figure 7, which has proven to be very effective in interpreting our XGBoost model. It allows us to visualize the impact of all explanatory variables in the XGBoost prediction model. Variable 5 (moving average of Δ_{Level} over 30 seconds), with a SHAP dispersion ranging from -4 to +8, is the most influential variable. High values in red contribute significantly to the increase in the variable Δ_{Level}_{t+10} . Low values in blue contribute to reducing this increase. Variables 6 (Δ_{Level}) and 7 (standard deviation over 30 seconds of Δ_{Level}) also contribute significantly, although less so than the first variable. The other variables have a limited contribution. We have chosen to keep the latter because they provide additional information that takes into account the dynamics of the system.

4. Conclusion

At the end of our study, it appears that the electrolyzer is a complex system, with multiple interactions between its various subsystems. We therefore modeled the 1 MW alkaline electrolyzer and then broke it down into subsystems. Our analyses highlighted the impact of controlling the liquid level control valve in the H_2 and O_2 separators. To this end, we developed machine learning models based on a set of explanatory variables that enable the valve to better

anticipate level deviations. All models performed well, with an R^2 score of 0.99 and an RMSE of less than 0.15 for Random Forest, XGBoost, DNN, ANN, and CatBoost. Linear regression, despite its simplicity, achieved a good score ($R^2 = 0.98$, RMSE = 0.40). Finally, the Prophet model had a less impressive score of $R^2 = 0.86$ and RMSE = 1.04. XGBoost was selected as the best model due to its robustness, its handling of non-linearity and, above all, its rapid inference capability ($R^2 = 0.99$ and RMSE = 0.10).

This study provides a solid foundation for future developments. The prediction model will be integrated into the control loop and the performance gains of the electrolyzer will be evaluated. The additional operating data collected will be added to the model, which will then be extended to the entire fleet of 1 MW electrolyzers.

Overall, our work is part of the digitalization of all installed electrolyzers and future projects, through the use of data for efficient production.

References

- [1] S. Iqbal, K. Aftab, F. tul Jannat, M. A. Baig, U. Kalsoom, A bibliographic analysis of optimization of hydrogen production via electrochemical method using machine learning, *Fuel* 372 (2024) 132126.
- [2] M. Motiramani, P. Solanki, V. Patel, T. Talreja, N. Patel, D. Chauhan, A. K. Singh, Ai-ml techniques for green hydrogen: A comprehensive review, *Next Energy* 8 (2025) 100252.
- [3] K. Adeli, M. Nachtane, A. Faik, A. Rachid, M. Tarfaoui, D. Saifaoui, A deep learning-enhanced framework for sustainable hydrogen production from solar and wind energy in the moroccan sahara: coastal regions focus, *Energy Conversion and Management* 302 (2024) 118084.
- [4] A. Jabbar, N. Pourmahmoud, M. A. A. Abdollahi, M. A. Rosen, Artificial intelligence-assisted optimization and multiphase analysis of polygon pem fuel cells, *International Journal of Green Energy* 21 (2024) 1550–1566.
- [5] A. Salari, H. Shakibi, M. A. Soleimanzade, M. Sadrzadeh, A. Hakkaki-Fard, Application of machine learning in evaluating and optimizing the hydrogen production performance of a solar-based electrolyzer system, *Renewable Energy* 220 (2024) 119626.
- [6] Z. Shangguan, H. Li, B. Yang, Z. Zhao, T. Wang, L. Jin, C. Zhang, Optimization of alkaline electrolyzer operation in renewable energy power systems: A universal modeling approach for enhanced hydrogen production efficiency and cost-effectiveness, *International Journal of Hydrogen Energy* 49 (2024) 943–954.
- [7] A. Sirat, S. Ahmad, I. Ahmad, N. Ahmed, M. Ahsan, Integrative cfd and ai/ml-based modeling for enhanced alkaline water electrolysis cell performance for hydrogen production, *International Journal of Hydrogen Energy* 83 (2024) 1120–1131.
- [8] K. E. Bassey, C. Ibegbulam, Machine learning for green hydrogen production, *Computer Science & IT Research Journal* 4 (2023) 368–385.
- [9] Z. Liu, Z. Cui, M. Wang, B. Liu, W. Tian, A machine learning proxy based multi-objective optimization method for low-carbon hydrogen production, *Journal of Cleaner Production* 445 (2024) 141377.
- [10] J. Lee, S. Hong, H. Cho, B. Lyu, M. Kim, J. Kim, I. Moon, Machine learning-based energy optimization for on-site smr hydrogen production, *Energy Conversion and Management* 244 (2021) 114438.
- [11] S. A. Bonab, T. Waite, W. Song, D. Flynn, M. Yazdani-Asrami, Machine learning-powered performance monitoring of proton exchange membrane water electrolyzers for enhancing green hydrogen production as a sustainable fuel for aviation industry, *Energy Reports* 12 (2024) 2270–2282.
- [12] G. Cheng, E. Luo, Y. Zhao, Y. Yang, B. Chen, Y. Cai, X. Wang, C. Dong, Analysis and prediction of green hydrogen production potential by photovoltaic-powered water electrolysis using machine learning in china, *Energy* 284 (2023) 129302.
- [13] Q. Yang, L. Zhao, J. Xiao, R. Wen, F. Zhang, D. Zhang, Machine learning-assisted prediction and optimization of solid oxide electrolysis cell for green hydrogen production, *Green Chemical Engineering* 6 (2025) 154–168.
- [14] J. Koponen, A. Kosonen, V. Ruuskanen, K. Huoman, M. Niemelä, J. Ahola, Control and energy efficiency of pem water electrolyzers in renewable energy systems, *International journal of hydrogen energy* 42 (2017) 29648–29660.
- [15] A. S. Emam, M. O. Hamdan, B. A. Abu-Nabah, E. Elnajjar, A review on recent trends, challenges, and innovations in alkaline water electrolysis, *International Journal of Hydrogen Energy* 64 (2024) 599–625.
- [16] Á. Hernández-Gómez, V. Ramirez, D. Guilbert, Investigation of pem electrolyzer modeling: Electrical domain, efficiency, and specific energy consumption, *International journal of hydrogen energy* 45 (2020) 14625–14639.