



**HAL**  
open science

## Time to Market Reduction for Hydrogen Fuel Cell Stacks using Generative Adversarial Networks

Nicolas Morizet, Perceval Desforges, Christophe Geissler, Elodie Pahon, Samir Jemeï, Daniel Hissel

► **To cite this version:**

Nicolas Morizet, Perceval Desforges, Christophe Geissler, Elodie Pahon, Samir Jemeï, et al.. Time to Market Reduction for Hydrogen Fuel Cell Stacks using Generative Adversarial Networks. 2022. hal-03906212

**HAL Id: hal-03906212**

**<https://hal.science/hal-03906212>**

Preprint submitted on 21 Dec 2022

**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.

# Time to Market Reduction for Hydrogen Fuel Cell Stacks using Generative Adversarial Networks

Nicolas Morizet<sup>a</sup>, Perceval Desforges<sup>a</sup>, Christophe Geissler<sup>a</sup>, Elodie Pahon<sup>b</sup>,  
Samir Jemei<sup>b</sup>, Daniel Hissel<sup>b</sup>

<sup>a</sup>*Advestis, 69 Boulevard Haussmann, Paris, 75008, France*

<sup>b</sup>*FEMTO-ST Institute, FCLAB, Univ. Bourgogne Franche-Comte,  
CNRS, Belfort, France*

---

## Abstract

To face the dependency on fossil fuels and limit carbon emissions, fuel cells are a very promising technology and appear to be a key candidate to tackle the increase of the energy demand and promote the energy transition. To meet future needs for both transport and stationary applications, the time to market of fuel cell stacks must be drastically reduced. Here, a new concept to shorten their development time by introducing a disruptive and high-efficiency data augmentation approach based on artificial intelligence is presented. Our results allow reducing the testing time before introducing a product on the market from a thousand to a few hours. The innovative concept proposed here can support engineering and research tasks during the fuel cell development process to achieve decreased development costs alongside a reduced time to market.

*Keywords:* Hydrogen fuel cell, Generative Adversarial Networks

*PACS:* 0000, 1111

*2000 MSC:* 0000, 1111

---

## 1. Introduction

In 2021, more than 20 countries had already published or were actively preparing their national strategies relating to hydrogen [1]. These strategies are commonly based on three pillars: hydrogen production [2, 3, 4], hydrogen storage and distribution [5, 6, 7], and hydrogen usages [8]. Regarding the latter, hydrogen can be used directly in chemical industry and in combustion engines, but also converted to electricity (and heat) through fuel cells

dedicated to stationary [9, 10] and transport applications [11, 12, 13, 14]. The global fuel cell market has been sharply increasing in the last couple of years [15] considering the very large range of possible applications. In this dynamic framework and to meet the market requirements as well as possible, two main research areas are still strongly investigated from the industrial and academic point of views: the increase of efficiency and durability of fuel cell stacks and systems together with the reduction of their cost [15, 16]. Obviously, in the V-model of a project aiming at developing a new fuel cell stack, testing and verification of both performances and durability of the developed product play a key role in terms of duration and costs. To reach these objectives, a disruptive solution involving generative adversarial networks (GAN) to decrease these fuel cells' test duration and validation costs is developed. GAN have been already used for the real-time smart generation control of microgrids [17, 18, 19, 20] as well as for the electricity prices forecasting [21] even by including data augmentation for time series regression [22]. It is also used for diagnostic tasks such as in [23], to detect failures on chiller systems to significantly reduce energy consumption and improve energy efficiency of buildings. GAN are also applied in the lithium-ion battery field for capacity estimation. GAN is utilized to obtain meaningful latent variables extracted from the impedance spectroscopy [24]. However, this approach has not yet been applied to fuel cells. However, it seems quite appropriate method to contribute significantly reducing the time to market for a given new generation of fuel cell stack.

## **2. Time to market for fuel cell systems - state of the art**

Without any reusable previous experiment, it is estimated that the development time for a new fuel cell system is 15 to 20 years [25]. This very long development time includes all stages, from the selection of the best materials for the catalyst to the integration into a complete system, including the development of bipolar plates and their industrialization. When considering integration of existing fuel cell stacks into systems by OEM (Original Equipment Manufacturer), this development time can be reduced to about 18 to 24 months. Nevertheless, in both situations, functional and performance testing of the fuel cell stack is required, mostly done according to international test protocols [26]. These tests, even excluding long duration tests, are highly costly and time-consuming. A full characterization of a stack in various operating conditions, as defined in these international test protocols, can be

estimated to a couple of weeks.

Moreover, focusing on industrial production, and according to an analysis made for the US Department of Energy in 2017 [27], total testing time per produced fuel cell stack is assumed to be about 2.5 hours. This testing and validation time includes the connection of the stack to the appropriate air, hydrogen, and cooling subsystems and to an appropriate controlled load bank. This figure must also be put in parallel with the time required to manufacture a whole fuel cell stack, estimated to 13 minutes for an automotive fuel cell stack in 2021 [28], leading to an end-of-line test time 10x greater than assembly time.

Once a stack is assembled and validated on the production line, the next step is to characterize its performance under different operating conditions, in order to qualify it for integration in various kind of power modules. The reference test is primarily the stack voltage obtained for different values of the drawn current (so-called polarization curves) under the applied test conditions. Polarization curves are often baseline measurements to qualify PEMFC (Proton Exchange Membrane Fuel Cell) stacks and components (e.g. catalyst, membrane, membrane electrode assembly, bipolar plates, etc) for given test conditions. To perform a polarization curve, a normalized test procedure has to be followed. It consists in different steps: start-up, conditioning, stabilization and measurements. Other recommended tests, also in different operating conditions, are the measurement of the voltages of the individual cells, the coolant outlet temperature (considered as the stack temperature), the outlet temperatures of the reactants (hydrogen and air) and the inlet and outlet pressures of the reactants.

With a rapidly expanding market, recent studies are looking at ways to further reduce the time required to break-in a fuel cell which is currently not suitable for large scale industrial applications. The time spent on the conditioning bench and the associated costs, mainly based on hydrogen consumption, must also be reduced [29].

### 3. Results

We have built and tested a data generator capable of producing artificial data in a short time that offer a high degree of statistical fit to the results of real experiments. These results come directly from the test bed, which produced data during five consecutive days of testing under normal operating conditions. The retained experimental design consists in the variation of 10

input variables (*i.e.*  $T_{out_{coolingwater}}$ ,  $T_{in_{H_2}}$ ,  $T_{in_{Air}}$ ,  $P_{in_{Air}}$ ,  $P_{in_{H_2}}$ ,  $Q_{in_{Air}}$ ,  $Q_{in_{H_2}}$ ,  $RH_{in_{Air}}$ ,  $RH_{in_{H_2}}$ ,  $Q_{coolingwater}$ ) across value ranges considered valid by the experimenter. 52 physical measurements including the previous 10 input variables, the main output variable representing the stack output voltage, and 41 other output variables (Table A.2) were recorded at a frequency of  $\approx 1$  Hz. In total, we have 30,901 observations corresponding to the stabilization and polarization steps of the fuel cell. Depending on the experiment, extra categorical features (*day* and *step name*) are added to the continuous ones.

### 3.1. Training

Model training is performed on stabilization and polarization data taken on a five days span (very short period compared to the expected lifetime of the fuel cell). Input continuous data are standardized and categorical variables are one-hot encoded before being fed to a Generative Adversarial Network (GAN). In this work, the data reproduced are not time series (an attempt to reproduce synthetic energy time series using recurrent GANs can be found in [30]), but synchronous snapshots of all the variables of interest from the stabilization and polarization phases. The measurements do not show any time dependency - besides the particular schedule retained for the physical inputs (temperature, pressure, current) in the experiments. Therefore, the choice made for this experiment was to validate some physical, non-temporal relationships between variables (e.g cell output voltage and stream intensity). The training ends after 25,000 epochs (each epoch corresponds to one batch of training data used to update the parameters of both networks once), when the critic can no longer improve its ability to distinguish artificial data from real data (their critic scores are very close), and the generator can no longer improve the quality of artificial data sent to the critic. On a local machine with a NVIDIA GeForce GTX 1080 Ti equipped with 12GB of video memory, the full training takes  $\approx 1$  hour. The architecture and the hyperparameters of the network are detailed in the *Methods* section.

### 3.2. Inference

Once the training is finished, the GAN model is completely described by  $\approx 218,000$  parameters. The generator is then detached and saved as a stand-alone function that can be called at will. This function takes as input a vector of 150 coordinates corresponding to independent draws of normalized Gaussian variables called the latent space. It is often agreed [31] that increasing the dimension of the latent space decreases the distortion of data,

at the cost of a polynomial increase in the computation time. The latent space is mapped onto the space of real variables by the non-linear generator function. The latter provides as output vectors of length 52 (as many entries as continuous variables to reproduce), which are supposed to mimic experimental data performed at different physical conditions. The generator function can be called on a machine with a much smaller configuration than the one used for training. This situation is typical of neural networks, which require large computational resources (RAM and GPU) in training phase, but much smaller resources in inference phase. For example, the detached generator can produce 100,000 vectors of 52 artificial data on a 12 cores CPU, in  $671.12\text{ms} \pm 0.024$  (statistics obtained for 500 runs), while requiring at most 2 GB of RAM.

### 3.3. Results qualification

The quality assessment of data produced by a GAN has been mostly studied in the field of image generation [32, 33, 34, 35]. Establishing the quality of artificial tabular data representing physical quantities requires different methods from those usually used for image processing. The qualitative role of the human eye cannot be directly transposed to the evaluation of physical tabular data. The verification of the closeness between the distributions of the real and generated data focuses on the quality of reproduction of the correlations between the variables. We introduce several useful metrics for data comparison.

### 3.4. Correlation matrices

The accordance of correlation matrices is the first test to pass for the evaluation of tabular data. The signs of these correlations are corroborated by physical relationships between the variables. Failure by the GAN to reproduce correlations of the same sign would signal failure to learn the structural relations linking the data. The visual representation of those matrices (Figure 1) indicates the absence of any sign error. All the variables are considered in the calculation of the correlations, but for visualization purpose, only a subset are shown: the ten input variables set by the experimental design along with the stack output voltage ( $V_{stack}$ ). The Kendall similarity score is used to test the similarities in the ordering of data when it is ranked by quantities [36]. It originally lies within  $[-1, 1]$  and is mapped in  $[0, 1]$ , the higher the better. In this first experiment, 100,000 data are generated and the Kendall similarity score is computed for the two matrices. The score is

0.95 (the p-value being less than 1%). For all the pairs of variables expected to show a significant correlation (e.g greater than 0.50 in absolute value), no sign error is found and the absolute difference is always less than 0.20.

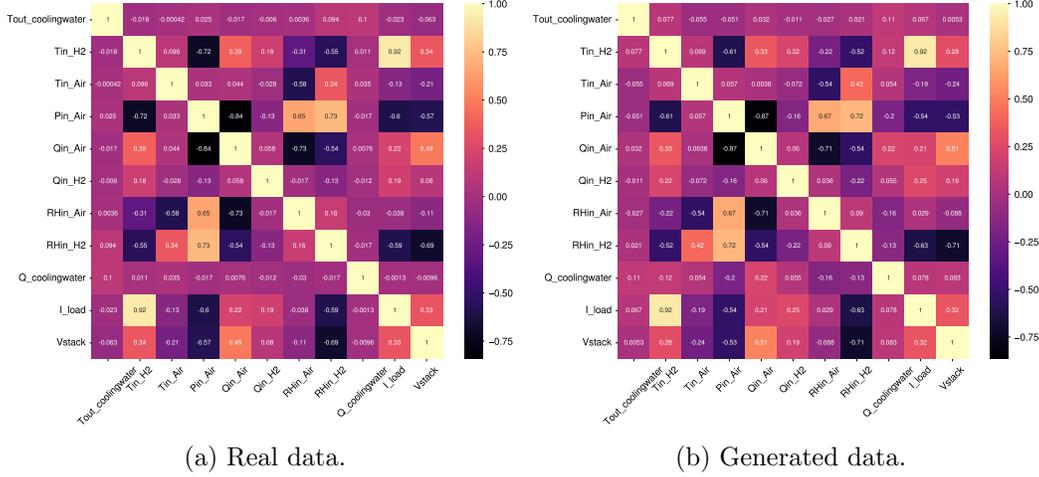


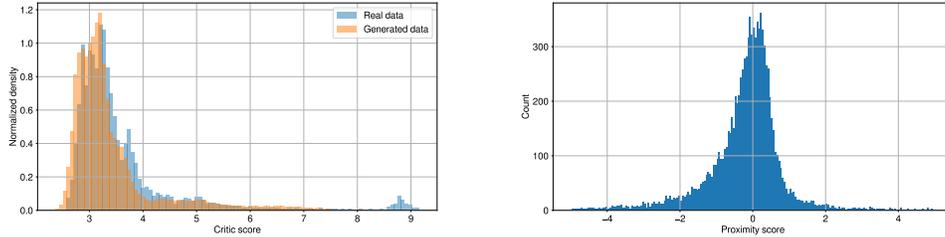
Figure 1: Correlation matrices of continuous features (stabilization step).

### 3.5. Proximity study

Another way to appreciate the closeness of the generated data to the real data is to use the critic scores assigned to them. The left part of Figure 2 depicts the overlay of the distributions of critic scores assigned to real and 100,000 generated data. Then, a random subset of 10,000 generated data is considered. For each of those vectors of artificial data, the  $k = 20$  closest real data are found in the training set using the L2 norm restricted to the set of experimental input variables mentioned in the previous section. Then, the GAN critic is called to score the artificial data row and its nearest neighbors selected from the real data. A statistic on the difference over the generated dataset defines a proximity score (the closer to 0 the better). Proximity score distribution is depicted in the right part of Figure 2.

### 3.6. Triangle plot

Unlike images, tabular data have no natural visual representation. The graphical representation of the bivariate distributions calculated on all pairs of variables nevertheless provides a visual tool for assessing the closeness



(a) Distributions of critic scores assigned to real and generated data. (b) Proximity score distribution.  $\approx 98\%$  of values are kept after removing several outliers. This operation leads to a cropped version of the original distribution within  $(-5, 5)$ , having a mean of  $\mu = -0.20$  and a standard deviation of  $\sigma = 1.02$ .

Figure 2: Proximity study results (stabilization and polarization steps).

between real and generated data. The triangle plot depicted in Figure 3 illustrates the replication of univariate modes and complex correlations between variables, and visually extends the comparison of correlation matrices. One can also appreciate the great similarity between the 68% and the 95% confidence levels as an indication of the good performance of the generative model.

### 3.7. Polarization curve

Among the characteristics that we wish to replicate, the most important are the polarization curves. Indeed, these are used routinely to test and troubleshoot fuel cell systems. In order to assess the quality of the reproduction of the polarization curve, a polarization *closeness error metric* is defined (see Section *Methods*). Figures 4 and 5 illustrate respectively a very good and poor replication of the polarization curve from the GAN. In Figure 5, the learning set has been filtered so as to only keep observations having the variable  $I_{load}$  within  $[110A, 120A]$ . One important finding of this experiment is about data distribution: the GAN can be successfully trained on a reduced set of data, as shown in the subsampling study (Table 1), as long as the training data are sufficiently well distributed in space. The GAN shows in this case an interesting capacity to interpolate between experimental points. Conversely, when the training is restricted to a small region of the space, the GAN fails to reproduce the data, even within the training area.

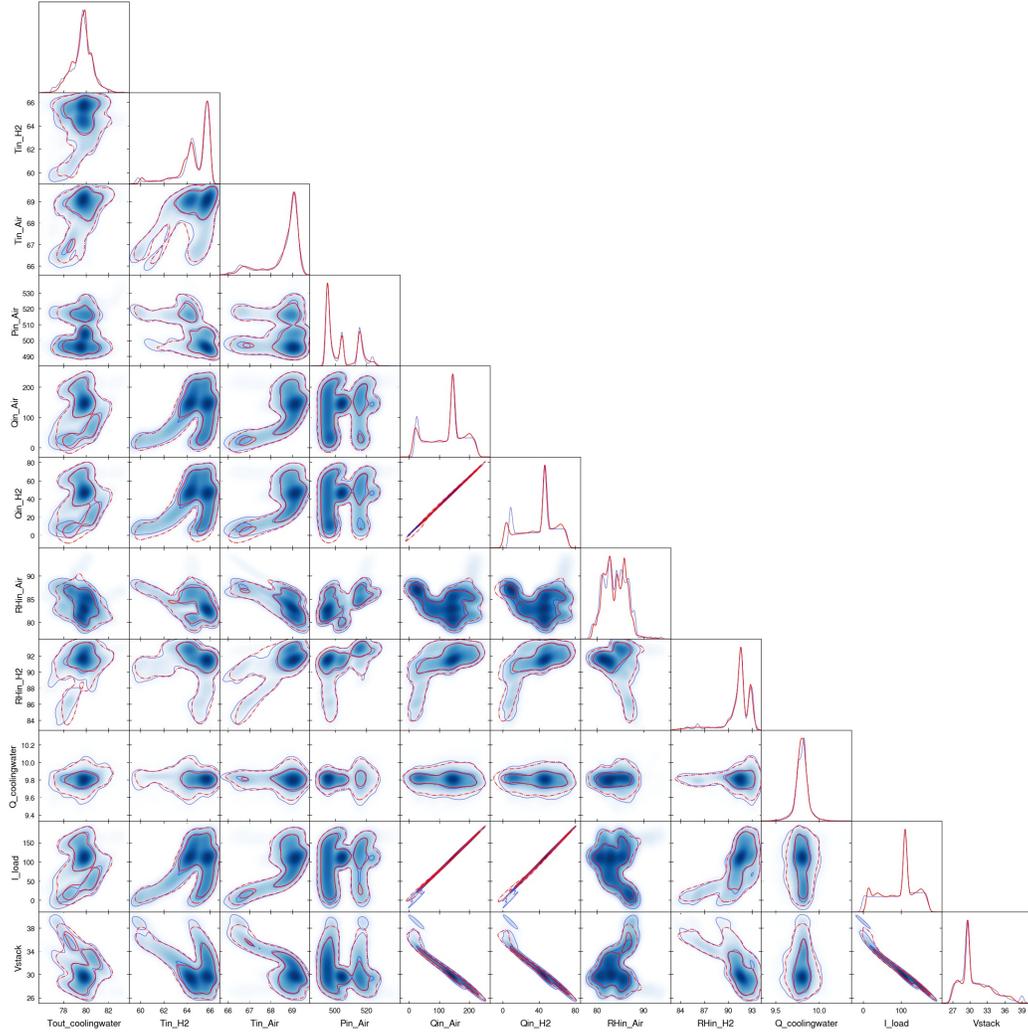


Figure 3: Triangle plot (stabilization and polarization steps): comparison between the real (*blue*) and generated (*red*) datasets via visual inspection of the respective marginal distributions. The one-dimensional (*diagonal panels*) and the two-dimensional (*off-diagonal panels*) marginal distributions are sorted by couples of continuous features. In each subplot, the 68% (*solid line*) and the 95% (*dash-dotted line*) confidence levels are also shown.

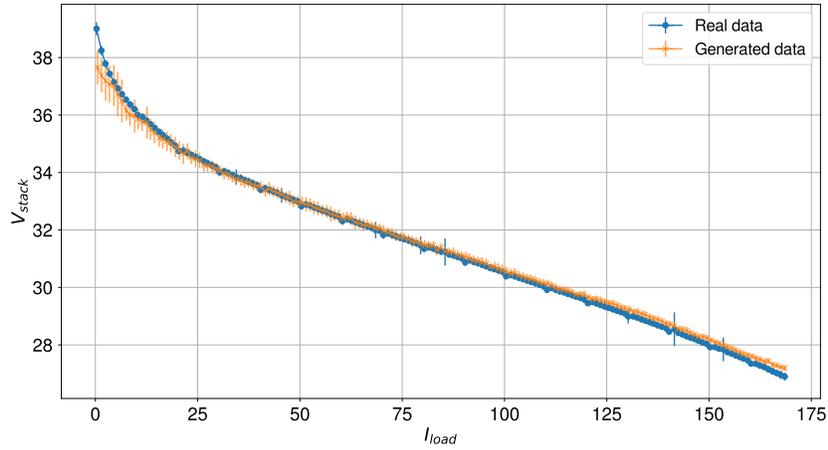


Figure 4: Polarization curve replication ( $\approx 25,000$  polarization data vs.  $\approx 25,000$  generated data). Polarization closeness errors are equal to 0.41% and 0.66%, for  $V_{stack}$  and  $I_{load}$ , respectively. These error rates are to be compared with the measurement errors for cell voltages and load currents which are less than 1%. The highest discrepancies appear at both ends of the range, but the non-linear relationship is overall well captured.

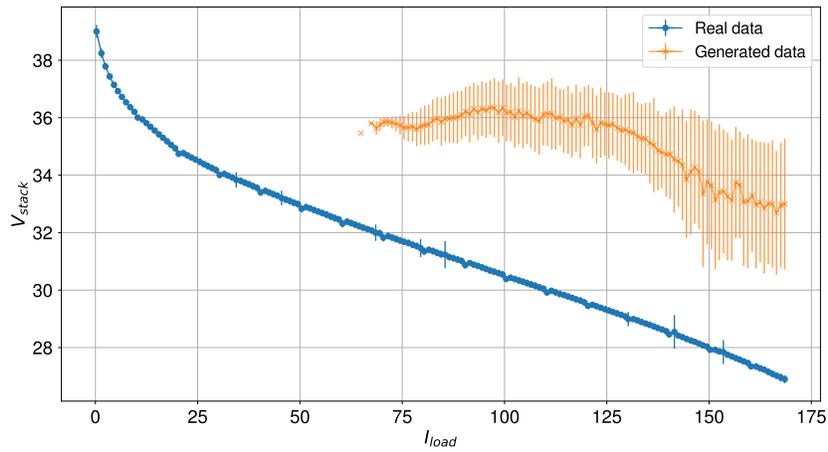


Figure 5: Polarization curve replication ( $\approx 25,000$  polarization data vs.  $\approx 25,000$  generated data). The same model has been trained only on data points having their load current variable in [110A, 120A]. This is an example of a bad sampling strategy, leading to a polarization closeness error of 18.8% for  $V_{stack}$ , way beyond the acceptable measurement error for cell voltages of 1%.

### 3.8. Subsampling study

The objective of this study is to analyze the impact of the number of real data available on the performance of the GAN. The real data are randomly and uniformly selected from the original dataset (preserving the original distribution for both classes – 20% for stabilization, 80% for polarization) and serve for training the GAN. This operation is equivalent to gradually lowering the original data acquisition factor, here from 1 down to 1/4. Performance metrics are shown in Table 1. What we can learn from this study is that, despite some statistical fluctuations, those performance metrics remain high, proving that the GAN does not need as much data as the original dataset (five days of experiments) to generate artificial data with a high level of similarity.

Data acquisition factor	$f = 1$	$f = 1/2$	$f = 1/4$
Real data (1)	30,901	15,450	7,725
Generated data (2)	100,000	100,000	100,000
Ratio (2) / (1)	3.24×	6.47×	12.9×
Epochs	25,000	25,000	25,000
Batch size	256	256	256
Training time (min)	60.2	60.0	60.8
KS score	0.87 ( $\pm 1.5e-3$ )	0.88 ( $\pm 1.5e-3$ )	0.87 ( $\pm 1.0e-3$ )
Dim red score	0.97 ( $\pm 4.1e-3$ )	0.91 ( $\pm 6.4e-3$ )	0.97 ( $\pm 1.7e-5$ )
Kendall score	0.90 ( $\pm 3.5e-4$ )	0.89 ( $\pm 1.9e-4$ )	0.90 ( $\pm 2.3e-4$ )
Pola error (%)	0.64 ( $\pm 2.6e-5$ )	0.44 ( $\pm 2.0e-5$ )	0.65 ( $\pm 2.8e-5$ )

Table 1: Performance results for the subsampling study on stabilization and polarization steps. In each experiment, one GAN model is trained and 100,000 data are generated and compared to the original dataset (with 30,901 real data). Performance metrics are given in terms of means and standard deviations over 100 different inferences. The Kendall score is mapped in the range [0, 1] and the polarization closeness error metric (see Section *Methods*) utilizes polarization data only.

## 4. Discussion

In fuel cell systems, many parameters must be set in specific ranges in order to achieve an optimal performance of the device, such as the reactants' temperatures, pressures, flows, and hydration levels, the stack temperature,

the current, and the voltage. One of the challenges of this study is to find relationships between the artificial variables that are dictated by the underlying physics. For instance, the gas flow is proportional to the current. The proposed GAN architecture produces artificial data that faithfully reproduces the joint distribution of the real ones. A high replication quality is achieved, as illustrated by the calculation of several statistical metrics. It is however important to underline the strong dependence of the resulting generator on the experimental design used to produce the data. The analysis of the joint distributions shows that for each pair of input variables, only a small fraction of the respective ranges of variation was explored. There are at least two reasons for this. The first is related to the number of acquisitions. If for every input variable their entire range of values must be explored, the total amount of acquisitions would increase exponentially which is clearly unfeasible. The second reason is physical and involves the experimenter’s know-how: not all combinations of variables are valid experimental conditions, some of them not allowing the stabilization of the stack, or even leading to physical degradations. The experimental design appears therefore far from arbitrary. If it is improperly chosen, like in the example presented for the Polarization Curve, the GAN may lose its ability to replicate data. The main strength of the current generator is its ability to interpolate data within the multi-dimensional experimental space, thus reducing the time spent by the fuel cells on the test bench.

As mentioned above, the bivariate plots show a visually very high degree of replication of the experimental design by the GAN (Figure 3). For example, the marginal distribution of the couple of variables ( $V_{\text{stack}}, I_{\text{load}}$ ) is captured with a very good precision. In order to be generalized in higher dimension, this visual fit can be translated into image similarity scores. In this particular case, the most important physical relationships involve at most two variables, which makes the verification of the trivariate marginal distributions less crucial, keeping in mind that this verification has cubic complexity with respect to the number of variables.

The polarization curve is replicated with a relative error that is, on average, less than the measurement uncertainty (Figure 4). A maximum distortion of 1.5% appears on the output voltage in the region of low currents (less than 5A). This distortion is very likely to be reduced by further improvements in the GAN architecture.

Considering the objective of reducing as much as possible the time spent on the test bench and the associated costs, it is possible to go from 4 hours

of measurements (relative to one polarization curve) to almost one hour. It is important to note that once the model is trained, it only takes less than 35 ms to generate a polarization curve with 5,000 points . This guarantees a considerable time saving when evaluating the performances of a fuel cell stack.

## 5. Conclusion

The fast generation of reliable and consistent artificial data paves the way to promising developments and operational improvements in fuel cell qualification. According to Table 1, the characterization time can be conservatively cut by 4. A next step will be the extension of the generator to non-nominal operating modes (e.g. air and fuel varying stoichiometries, high and low operating temperatures) to ensure the ability to map the behavior of the fuel cell stack in all operating conditions. The expected benefits of data generation are even greater in default mode than in normal operation, since experiments under fault conditions degrade the equipment and can therefore only be performed a limited number of times. A further step will be to use these artificial data collected in normal and default modes to train a model of the fuel cell performance throughout its lifetime.

## 6. Methods

### 6.1. Data from experimental tests

We consider a proton exchange membrane fuel cell (PEMFC) with an active area of  $220\text{cm}^2$ , a nominal operating temperature of  $80^\circ\text{C}$ , and a nominal current density of  $0.5\text{A}/\text{cm}^2$ . It is composed of 40 cells and delivers up to 3kW of electric power. On the test bench, various parameters are supervised and regulated such as the pressures, the flow rates, the gas humidity, and the dew points. The data generation is based on several experimental tests done on the fuel cell stack, in various operating conditions. First, an experiment under nominal operating conditions is performed. To evaluate the performances of the fuel cell stack, a polarization curve (voltage vs. current curve) is performed. Then, it is required to investigate the behavior of the fuel cell stack under degraded operating conditions. Thus, the values of the parameters are progressively increased or decreased up to a given threshold, defined by the physical limits of the test bench and by the fuel cell specifications. The full generated dataset is thus divided into several parts: *i*)

start-up of the fuel cell, *ii*) stabilization of the fuel cell at nominal operating conditions, *iii*) polarization curves and *iv*) shutdown. The start-up time of the fuel cell depends on its subsequent use. Indeed, if the fuel cell is brand new, it will be necessary to break it in. Even if some optimization methods can be applied to the fuel cell manufacturing process to reduce this time, it will take multiple hours. If the time to perform a polarization curve is added to the "break-in" process then this will further increase the time on the test bench. To consider a mass production of fuel cells at low cost, it is mandatory to reduce the time spent per fuel cell on a test bench or to use the concept presented in this paper. After data merging, this currently represents approximately 4 hours.

## 6.2. GANs architecture

Generative Adversarial Networks (GANs) were first introduced[37] in 2014 by Ian Goodfellow. They belong to the family of generative models and are still considered to be the state of the art for approximating real data distributions. They allow to sample multivariate probability distributions with implicit density. GANs involve two actors: one is called the generator, the other the discriminator. The goal of the generator is to create fake samples that are supposed to come from the same distribution as the original training data. The discriminator tries to determine whether its input data are real or fake. The competition in this game pushes both entities to improve themselves until the fake data are indistinguishable from the real ones. This sweet spot is called the Nash equilibrium[38] and corresponds to the moment when the model has completed its training. In practice, the generator and the discriminator are represented by differentiable functions such as a deep neural networks. In this work, we adopt as a blueprint a Wasserstein-GAN with gradient penalty (WGAN-GP)[39] (the discriminator is call the critic) in which we customize the architecture to handle continuous and categorical variables simultaneously. The WGAN-GP brings strong modeling performance and more stability for training compared to its predecessors. Moreover, it has a loss function based on the concept of optimal transport [40, 41], well suited to approximate the shape of statistical distributions which correlates with generated samples quality. Detailed structure of both the generator and the critic are available in Table B.3 and Table B.4 from the appendix, respectively. Regarding the generator architecture, The sub-module dedicated to the processing of a categorical variable is kept generic, indicating with  $k_i$  the number of classes for the  $i^{\text{th}}$  categorical variable. At training time, the out-

put layer of the generator is a concatenation of the layers B and  $C_{i=1,\dots,N}$ , for a total dimension of  $D = d + \sum_{i=1}^N k_i$ ,  $d$  being the number of non-categorical variables. The abbreviation *Conn.* stands for *Connected*. As far as the critic architecture is concerned, the abbreviation *dr.* stands for *dropout rate*.

### 6.3. GANs training parameters

The model is trained for 25,000 epochs. It was verified that without adding any further complexity to the overall architecture, pushing the training for more epochs was useless and could even lead to worse performance results. A batch size of  $n_b = 256$  data is chosen as a good compromise between rapidity and quality of the training. Adam optimizer [42] is used with a learning rate of  $l_r = 0.001$  for both the generator and the critic. The dimension of the latent space is set to  $L_{\text{dim}} = 150$  which, even if it is higher than the number of variables, helps to improve performance results by decreasing the distortion of the generated data, as mentioned in [31]. The gradient penalty is set to  $\lambda = 10$ , in line with previous literature [42] while the critic / generator training ratio is equal to 15. The total number of parameters of the model is 218,418 (120,769 for the generator and 97,649 for the critic).

### 6.4. Kolmogorov-Smirnov score

The two-sample Kolmogorov–Smirnov (KS) test [43] is used to verify whether two one-dimensional samples come from the same distribution. This metric compares the distributions of continuous features using the empirical cumulative distribution function (CDF). For each feature, the similarity score is computed as one minus the KS test D-statistic, which indicates the maximum distance between the real data CDF ( $CDF_{\text{real}}$ ) and the generated data CDF ( $CDF_{\text{gen}}$ ) values. The output score  $S_{\text{ks}}$  lies in  $[0, 1]$  (the higher the better) and is computed as the minimum score across all the features:

$$S_{\text{ks}} = \min_i (1 - \sup_{x_i} |CDF_{\text{real}}(x_i) - CDF_{\text{gen}}(x_i)|) \quad (1)$$

The minimum is taken instead of the mean that would smooth errors. Thus, if this score is low, we would have the information that at least one feature has not been approximated correctly. This metric quantitatively complements the visual inspection of the triangle plot (Figure 3) to assess the quality of the reproduction of marginal distributions.

### 6.5. Dimension reduction score

The dimension reduction score aims to reflect how well the model has captured the correlations between the different features. First, a Principal Component Analysis (PCA) [44] is performed on both real and generated data such that 99% of the variance is explained by the first  $n$  eigenvectors. Then, real and generated eigenvectors are normalized by the square root of their corresponding eigenvalues (resp.  $m_{\text{real}}$  and  $m_{\text{gen}}$ ) before a cosine similarity  $S_{\text{cosine}}$  is computed between them. The output score  $S_{\text{dim}}$  lies within  $[0, 1]$  (the higher the better) and is computed as follows:

$$S_{\text{dim}} = \frac{\sum_{i=1}^n \omega_i \times S_{\text{cosine}}^i(m_{\text{real}}^i, m_{\text{gen}}^i)}{\sum_{i=1}^n \omega_i}, i = 1 \dots n \quad (2)$$

with

$$S_{\text{cosine}}(x, y) = \frac{1}{2} \left( \frac{x \cdot y}{|x| \times |y|} + 1 \right) \in [0, 1] \quad (3)$$

where  $\omega_i$  corresponds to the explained variance ratio of the  $i$ th real eigenvector.

### 6.6. Proximity score

In the neighborhood study, the critic score from one generated row  $S_c^{\text{gen}}$  is compared to the distribution of the critic scores of its  $k$ -nearest neighbors  $S_c^{\text{knn}}$  using a proximity score  $S_{\text{prox}}^g$  computed as follows:

$$S_{\text{prox}}^g = \frac{S_c^{\text{gen}} - \min \{S_c^{\text{knn}}\}}{\max \{S_c^{\text{knn}}\} - \min \{S_c^{\text{knn}}\}} \quad (4)$$

### 6.7. Polarization closeness error

To compute the polarization closeness error, we divide the currents in  $nc = \max([I_{\text{load}}])$  bins, and compute the average of the currents and of the voltages for both real and generated data in each bin (which we will note  $\bar{V}_{\text{real},i}$ ,  $\bar{V}_{\text{gen},i}$ ,  $\bar{I}_{\text{real},i}$ , and  $\bar{I}_{\text{gen},i}$ ). This allows us to produce the box plots visible in Figure 4. We then take the average of the relative error between

real and generated data in each bin, which gives us the average percentage error  $e$  between both data types :

$$e = \frac{1}{nc} \sum_{i=1}^{nc} \left| \frac{\bar{V}_{\text{real},i} - \bar{V}_{\text{gen},i}}{\bar{V}_{\text{real},i}} \right| \quad (5)$$

### **Acknowledgements**

This work has been supported by the EIPHI Graduate School (contract ANR-17-EURE-0002) and the Region Bourgogne Franche-Comté

### **Appendix A. Experimental Data Information**

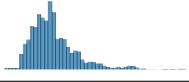
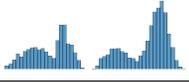
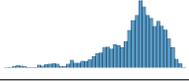
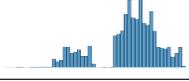
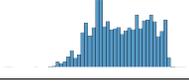
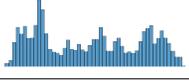
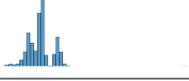
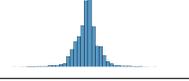
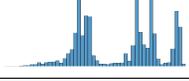
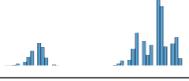
Feature name	Description	Unit	Range	Histogram
$T_{out_{coolingwater}}$	Outlet temperature of the water	°C	[25 - 95]	
$T_{in_{H_2}}$	Inlet temperature of the hydrogen	°C	[25 - 95]	
$T_{in_{Air}}$	Inlet temperature of the air	°C	[25 - 95]	
$P_{in_{Air}}$	Inlet pressure of the air	mbarg	[0 - 1000]	
$P_{in_{H_2}}$	Inlet pressure of the hydrogen	mbarg	[0 - 1000]	
$Q_{in_{Air}}$	Total inlet air flow rate	Nl/min	[0 - 300]	
$Q_{in_{H_2}}$	Total inlet hydrogen flow rate	Nl/min	[0 - 100]	
$RH_{in_{Air}}$	Relative humidity of the air (inlet of the stack)	%	[0 - 100]	
$RH_{in_{H_2}}$	Relative humidity of the hydrogen (inlet of the stack)	%	[0 - 100]	
$Q_{coolingwater}$	Deionized water flow rate	l/min	[0 - 15]	
$V_{stack}$	Stack voltage	V	[0 - 40]	
$I_{load}$	Current load	A	[0 - 200]	
$V_{cell_i}$	Cell voltage of the cell # $i$ ( $i = 1 \dots 40$ )	V	[0 - 1]	

Table A.2: Description of the continuous real data involved in the PEMFC experiments (stabilization step).

## Appendix B. GANs Architecture

Layer type	Output shape	Parameters	Info.
Input sub-module			
Input	$(n_b, 150)$	0	
Dense	$(n_b, 256)$	38656	
Batch Norm.	$(n_b, 256)$	1024	
$(s = +1)$ -SLR	$(n_b, 256)$	2	
Dense	$(n_b, 128)$	32896	
Batch Norm.	$(n_b, 128)$	512	
$(s = -1)$ -SLR	$(n_b, 128)$	2	
Dense	$(n_b, 64)$	8256	
Batch Norm.	$(n_b, 64)$	256	
$(s = +1)$ -SLR	$(n_b, 64)$	2	
Dense	$(n_b, 32)$	2080	
Batch Norm.	$(n_b, 32)$	128	
$(s = -1)$ -SLR	$(n_b, 32)$	2	Layer A
$d$ continuous variables			
Dense	$(n_b, 4d)$	$4d \cdot 33$	Conn. to A
Batch Norm.	$(n_b, 4d)$	$4d \cdot 4$	
$(s = -1)$ -SLR	$(n_b, 4d)$	2	
Dense	$(n_b, 2d)$	$(4d + 1) \cdot 2d$	
Batch Norm.	$(n_b, 2d)$	$2d \cdot 4$	
$(s = +1)$ -SLR	$(n_b, 2d)$	2	
Dense	$(n_b, d)$	$(2d + 1) \cdot d$	Layer B
$i^{\text{th}}$ categorical variable of $k_i$ classes ( $i = 1, \dots, N$ )			
Dense	$(n_b, 4k_i)$	$4k_i \cdot 33$	Conn. to A
Batch Norm.	$(n_b, 4k_i)$	$4k_i \cdot 4$	
LeakyReLU	$(n_b, 4k_i)$	0	
Dense	$(n_b, 2k_i)$	$(4k_i + 1) \cdot 2k_i$	
Batch Norm.	$(n_b, 2k_i)$	$(4k_i + 1) \cdot 4$	
LeakyReLU	$(n_b, 2k_i)$	0	
Dense	$(n_b, k_i)$	$(2k_i + 1) \cdot k_i$	
Batch Norm.	$(n_b, k_i)$	$(2k_i + 1) \cdot 4$	
Softmax	$(n_b, k_i)$	0	Layer $C_i$
Concatenate	$(n_b, D)$	0	Output (B, $C_{i=1, \dots, N}$ )

Table B.3: Generator architecture as function of the batch size  $n_b$ .

Layer type	Output shape	Parameters	Info.
Input	$(n_b, 59)$	0	
Dense	$(n_b, 256)$	15360	
LeakyReLU	$(n_b, 256)$	0	
Dropout	$(n_b, 256)$	0	No
Dense	$(n_b, 128)$	32896	
LeakyReLU	$(n_b, 128)$	0	
Dropout	$(n_b, 128)$	0	No
Dense	$(n_b, 128)$	16512	
LeakyReLU	$(n_b, 128)$	0	
Dropout	$(n_b, 128)$	0	No
Dense	$(n_b, 128)$	16512	
LeakyReLU	$(n_b, 128)$	0	
Dropout	$(n_b, 128)$	0	dr=0.5
Dense	$(n_b, 64)$	8256	
LeakyReLU	$(n_b, 64)$	0	
Dropout	$(n_b, 64)$	0	dr=0.5
Dense	$(n_b, 64)$	4160	
LeakyReLU	$(n_b, 64)$	0	
Dropout	$(n_b, 64)$	0	dr=0.2
Dense	$(n_b, 32)$	2080	
LeakyReLU	$(n_b, 32)$	0	
Dropout	$(n_b, 32)$	0	dr=0.2
Dense	$(n_b, 32)$	1056	
LeakyReLU	$(n_b, 32)$	0	
Dropout	$(n_b, 32)$	0	No
Dense	$(n_b, 16)$	528	
LeakyReLU	$(n_b, 16)$	0	
Dropout	$(n_b, 16)$	0	No
Dense	$(n_b, 16)$	272	
LeakyReLU	$(n_b, 16)$	0	
Dropout	$(n_b, 16)$	0	No
Dense	$(n_b, 1)$	17	Output

Table B.4: Critic architecture.

## References

- [1] IEA, Annual announcements of government hydrogen strategies, <https://www.iea.org/data-and-statistics/charts/annual-announcements-of-government-hydrogen-strategies-2018-september-2021> (2021).
- [2] J. Serra, J. Borrás-Morell, B. García-Baños, et al., Hydrogen production via microwave-induced water splitting at low temperature, *Nature Energy* 5 (1) (2020) 910–919. doi:<https://doi.org/10.1038/s41560-020-00720-6>.
- [3] H. Malerød-Fjeld, D. Clark, I. Yuste-Tirados, et al., Thermo-electrochemical production of compressed hydrogen from methane with near-zero energy loss, *Nature Energy* 2 (1) (2017) 923–931. doi:<https://doi.org/10.1038/s41560-017-0029-4>.
- [4] G. Glenk, S. Reichelstein, Economics of converting renewable power to hydrogen, *Nature Energy* 4 (1) (2019) 216–222. doi:<https://doi.org/10.1038/s41560-019-0326-1>.
- [5] J. Gallagher, Hydrogen storage: Cold feat, *Nature Energy* 1 (1) (2016) 16167. doi:<https://doi.org/10.1038/nenergy.2016.167>.
- [6] J. Gallagher, Hydrogen storage: Pressure swing, *Nature Energy* 2 (1) (2017) 17122. doi:<https://doi.org/10.1038/nenergy.2017.122>.
- [7] C. Zhang, Hydrogen storage: Improving reversibility, *Nature Energy* 1 (2) (2017) 17064. doi:<https://doi.org/10.1038/nenergy.2017.64>.
- [8] Z. Xia, Hydrogen evolution: Guiding principles, *Nature Energy* 1 (1) (2016) 16155. doi:<https://doi.org/10.1038/nenergy.2016.155>.
- [9] S. Pellegrino, A. Lanzini, P. Leone, Techno-economic and policy requirements for the market-entry of the fuel cell micro-chp system in the residential sector, *Applied Energy* 143 (2015) 370–382. doi:[10.1016/j.apenergy.2015.01.007](https://doi.org/10.1016/j.apenergy.2015.01.007).
- [10] M. Wei, S. Sarah, Smith, M. Sohn, Experience curve development and cost reduction disaggregation for fuel cell markets in japan and the us, *Applied Energy* 191 (2017) 346–357. doi:[10.1016/j.apenergy.2017.01.056](https://doi.org/10.1016/j.apenergy.2017.01.056).

- [11] Editorial, Hydrogen on the rise, *Nature Energy* 1 (2016) 16127. doi:<https://doi.org/10.1038/nenergy.2016.127>.
- [12] D. Cullen, K. Neyerlin, R. Ahluwalia, et al., New roads and challenges for fuel cells in heavy-duty transportation, *Nature Energy* 6 (1) (2021) 462–474. doi:<https://doi.org/10.1038/s41560-021-00775-z>.
- [13] C. Zhang, J. Greenblatt, M. Wei, J. Eichman, S. Saxena, M. Muratori, O. Guerra, Flexible grid-based electrolysis hydrogen production for fuel cell vehicles reduces costs and greenhouse gas emissions, *Applied Energy* 278 (2020) 115651. doi:[10.1016/j.apenergy.2020.115651](https://doi.org/10.1016/j.apenergy.2020.115651).
- [14] I. Kim, J. Kim, J. Lee, Dynamic analysis of well-to-wheel electric and hydrogen vehicles greenhouse gas emissions: Focusing on consumer preferences and power mix changes in south korea, *Applied Energy* 260 (2020) 114281. doi:[10.1016/j.apenergy.2020.114281](https://doi.org/10.1016/j.apenergy.2020.114281).
- [15] Editorial, Moving forward with fuel cells, *Nature Energy* 6 (2021) 451. doi:<https://doi.org/10.1038/s41560-021-00846-1>.
- [16] J. Palencia, Y. Otsuka, M. Araki, S. Shiga, Scenario analysis of lightweight and electric-drive vehicle market penetration in the long-term and impact on the light-duty vehicle fleet, *Applied Energy* 204 (2017) 1444–1462. doi:[10.1016/j.apenergy.2017.05.054](https://doi.org/10.1016/j.apenergy.2017.05.054).
- [17] K. Han, K. Yang, L. Yin, Lightweight actor-critic generative adversarial networks for real-time smart generation control of microgrids, *Applied Energy* 317 (2022) 119163. doi:[10.1016/j.apenergy.2022.119163](https://doi.org/10.1016/j.apenergy.2022.119163).
- [18] L. Yin, B. Zhang, Time series generative adversarial network controller for long-term smart generation control of microgrids, *Applied Energy* 281 (2021) 116069. doi:[10.1016/j.apenergy.2020.116069](https://doi.org/10.1016/j.apenergy.2020.116069).
- [19] J. Dumas, A. Wehenkel, D. Lanaspeze, B. Cornelusse, A. Sutera, A deep generative model for probabilistic energy forecasting in power systems: normalizing flows, *Applied Energy* 305 (2022) 117871. doi:[10.1016/j.apenergy.2021.117871](https://doi.org/10.1016/j.apenergy.2021.117871).
- [20] W. Dong, X. Chen, Q. Yang, Data-driven scenario generation of renewable energy production based on controllable generative adversarial networks with interpretability, *Applied Energy* 308 (2022) 118387. doi:[10.1016/j.apenergy.2022.118387](https://doi.org/10.1016/j.apenergy.2022.118387).

- [21] X. Lu, J. Qiu, G. Lei, J. Zhu, Scenarios modelling for forecasting day-ahead electricity prices: Case studies in australia, *Applied Energy* 308 (2022) 118296. doi:10.1016/j.apenergy.2021.118296.
- [22] S. Demir, K. Mincev, K. Kok, N. Paterakis, Data augmentation for time series regression: Applying transformations, autoencoders and adversarial networks to electricity price forecasting, *Applied Energy* 304 (2021) 117695. doi:10.1016/j.apenergy.2021.117695.
- [23] B. Li, F. Cheng, X. Zhang, C. Cui, W. Cai, A novel semi-supervised data-driven method for chiller fault diagnosis with unlabeled data, *Applied Energy* 285 (2021) 116459. doi:10.1016/j.apenergy.2021.116459.
- [24] S. Kim, Y. Choi, J. Choi, Impedance-based capacity estimation for lithium-ion batteries using generative adversarial network, *Applied Energy* 308 (2022) 118317. doi:10.1016/j.apenergy.2021.118317.
- [25] Lsu innovation in fuel cell technology for electric cars and trucks draws interest from General Motors, brings Toyota to campus, <https://www.lsu.edu/research/news/2021/0120-fuelcells.php>, march 3, 2022.
- [26] J. Hunger, et al., Development of pem fuel cell stack reference test procedures for industry, <http://stacktest.zsw-bw.de>, fCH JU Grant Agreement 303445 (2015).
- [27] B. M. Institute, Manufacturing cost analysis of 100 and 250kW fuel cell systems for primary power and combined heat and power applications, <https://www.energy.gov/eere/fuelcells/downloads/manufacturing-cost-analysis-100-and-250-kw-fuel-cell-systems-primary-0> (2017).
- [28] Mass production of fuel cells becomes a reality, [https://www.ipa.fraunhofer.de/en/press-media/press/\\_releases/mass-production-of-fuel-cell-becomes-a-reality.html](https://www.ipa.fraunhofer.de/en/press-media/press/_releases/mass-production-of-fuel-cell-becomes-a-reality.html), october 18, 2021.
- [29] E. Balogun, A. Oyarce, B. Holdcroft, Cathode starvation as an accelerated conditioning procedure for perfluorosulfonic acid ionomer fuel cells, *Journal of Power Sources Advances* 3 (2020) 100012. doi:https://doi.org/10.1016/j.powera.2020.1000120.

- [30] M. N. Fekri, A. M. Ghosh, K. Grolinger, Generating energy data for machine learning with recurrent generative adversarial networks, *Energies* 13 (1) (2020). doi:10.3390/en13010130.  
URL <https://www.mdpi.com/1996-1073/13/1/130>
- [31] T. Pinetz, D. Soukup, T. Pock, Impact of the latent space on the ability of GANs to fit the distribution (2020).  
URL <https://openreview.net/forum?id=Hygy01StvH>
- [32] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans (2016). arXiv:1606.03498.
- [33] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium (2018). arXiv:1706.08500.
- [34] K. Shmelkov, C. Schmid, K. Alahari, How good is my GAN?, *CoRR* abs/1807.09499 (2018). arXiv:1807.09499.  
URL <http://arxiv.org/abs/1807.09499>
- [35] Z. Wang, G. Healy, A. F. Smeaton, T. E. Ward, Use of neural signals to evaluate the quality of generative adversarial network performance in facial image generation, *Cognitive Computation* 12 (1) (2019) 13–24. doi:10.1007/s12559-019-09670-y.  
URL <http://dx.doi.org/10.1007/s12559-019-09670-y>
- [36] L. Puka, Kendall’s Tau, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 713–715. doi:10.1007/978-3-642-04898-2\_324.  
URL [https://doi.org/10.1007/978-3-642-04898-2\\_324](https://doi.org/10.1007/978-3-642-04898-2_324)
- [37] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks (2014). arXiv:1406.2661.
- [38] J. F. Nash, Equilibrium points in n-person games, *Proceedings of the National Academy of Sciences* 36 (1) (1950) 48–49. arXiv:<https://www.pnas.org/content/36/1/48.full.pdf>, doi:10.1073/pnas.36.1.48.  
URL <https://www.pnas.org/content/36/1/48>

- [39] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville, Improved training of wasserstein gans (2017). arXiv:1704.00028.
- [40] C. Villani, A. M. Society, Topics in Optimal Transportation, Graduate studies in mathematics, American Mathematical Society, 2003.  
URL <https://books.google.fr/books?id=MyPjjgEACAAJ>
- [41] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, G. K. Rohde, Optimal mass transport: Signal processing and machine-learning applications, IEEE Signal Processing Magazine 34 (4) (2017) 43–59. doi:10.1109/MSP.2017.2695801.
- [42] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization (2017). arXiv:1412.6980.
- [43] J. L. Hodges, The significance probability of the smirnov two-sample test, Arkiv for Matematik 3 (5) (1958) 469 – 486. doi:10.1007/BF02589501.  
URL <https://doi.org/10.1007/BF02589501>
- [44] Karl Pearson F.R.S., LIII. On lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2 (11) (1901) 559–572. doi:10.1080/14786440109462720.