

# A hybrid control oriented PEMFC model based on Echo State Networks and Gaussian Radial Basis Functions

José Agustín Aguilar <sup>\*,a</sup>, Damien Chanal <sup>b</sup>, Didier Chamagne <sup>b</sup>,

Nadia Yousfi Steiner <sup>b</sup>, Marie-Cécile Pera <sup>b</sup>, Attila Husar <sup>a,c</sup>, Juan Andrade-Cetto<sup>a</sup>

<sup>a</sup> Institut de Robòtica i Informàtica Industrial, CSIC-UPC

Llorens Artigas 4-6, 08028 Barcelona, Spain.

<sup>b</sup> Université de Franche-Comté, UTBM CNRS, institut FEMTO-ST, F-90000 Belfort, France

<sup>c</sup> Departament de Mecànica de Fluids, UPC

Pavelló 1, Diagonal 647, 08028, Barcelona, Spain.

## Abstract

The goal of increasing efficiency and durability of fuel cells can be achieved through optimal control of their operating conditions. In order to implement such controllers, accurate and computationally efficient fuel cell models must be developed. This work presents a hybrid, control-oriented model for proton exchange membrane fuel cells (PEMFC) operating under dynamical conditions. The model is formed by a parallel structure where a simplified model, built from electrochemical and mass conservation laws, is combined with an Echo State Network (ESN). A merging algorithm based on Gaussian radial basis functions compares the output of both the physics-based and the data-driven model and assigns weights for computing the prediction of the hybrid model. The proposed model structure is successfully trained, validated and tested with an experimental dataset originated from fuel cells within an automotive PEMFC stack.

**Keywords-** PEMFC, hybrid model, ESN, radial basis functions.

## 1 Introduction

The evidence of climate change induced by human activity has pushed the search for sustainable forms of production, storage and conversion of energy. Hydrogen has emerged as a promising energy carrier in a future low carbon economy, and fuel cells are likely the ideal option for converting hydrogen's chemical energy into electrical energy. In this context, transportation is among the sectors that require more

---

\*Corresponding author jaguilar@iri.upc.edu

aggressive decarbonization efforts and proton exchange membrane fuel cells (PEMFC) offer a real alternative to internal combustion engines in the powertrain of long distance, heavy load vehicles. However, in order to consider PEMFCs a viable commercial option, issues such as cell durability and efficiency must be addressed, specifically, the durability target has been set in 25,000 hours for those applications [1]. This performance is still out of reach for commercially available PEMFCs.

The goal of increasing efficiency and durability of PEMFC can only be accomplished through proper control of its operating conditions. Several studies have shown that operating conditions have a major influence on fuel cell integrity and efficiency. Some studies have identified deterministic relations between operating conditions, such as voltage profile or cell temperature, and the dissolution of platinum in the catalyst layer [2]; modeled the effect of voltage profile and humidity on the corrosion of the catalyst layer carbon support [3]; validated, experimentally, expressions for platinum dissolution, platinum oxidation and more chemical degradation mechanisms as function of cell temperature, humidity and load profile [4]; or related, through semi-empirical models, the degradation of the catalyst layer to the operating conditions in automotive applications [5].

Therefore, improvement in fuel cell durability and efficiency depends on optimal control of the operating conditions and for such control, accurate and computationally efficient models are required. Due to the complex and nonlinear interactions between its constitutive physical parameters, internal states and external operating conditions, fuel cells are challenging systems for modelling tasks.

Some models combine static and dynamic fuel cell behavior for prognosis purposes [6]. Also, in order to simplify the high non-linearity of fuel cell analytical expressions there are linearized parameter varying models [7]. PEMFC dynamics have also been modeled using equivalent electric circuits [8]. There has been extensive research on developing physics-based models of PEMFC and its auxiliary systems, with several degrees of accuracy and computational cost [9].

In general, there is an increasing interest on complementing models derived from the knowledge of underlying physical phenomena, with models built exclusively from data. The goal of those hybrid architectures is to take advantage of the respective strength in the physic-based and data-driven approaches, with the objective of representing complex behaviors, non-linearities, non stationary dynamics or time varying unknown parameters. Hybrid modelling of complex nonlinear systems have been proposed in areas such as manufacturing [10]; food industry [11]; chemical/oil/gas industry [12], [13], [14]; or biological systems [15], [16], [17].

In research areas closer to fuel cell research, hybrid approaches have been proposed for estimating the remaining useful life of lithium ion batteries through the merge of a Kalman filter and a neural network [18]. In that structure, the Kalman filter estimates the hidden internal states while the neural network serves as the observation expression, forecasting future values of the observation variable. Also, a parallel structure of neural network and Kalman filter have been proposed for state estimation and

state forecasting. [19].

The present work proposes a hybrid, control oriented model to forecast the output voltage of a PEMFC operating under dynamical conditions. The aim of this paper is to fill the research gap in computationally efficient PEMFC models that able to represent behaviors of fuel cells that are hard to forecast and that appear when the operating conditions vary dynamically. The main contribution of this work is the development, implementation and testing of a hybrid modelling structure formed by three main components: a simplified physics-based model, an Echo State Network (ESN) and a merging algorithm based on Gaussian radial basis functions (GRBF). The novel characteristics of the present work are detailed as follows:

1. The dataset used to train, test and validate the hybrid model comes from experiments where the operating conditions and thus the profile of the output voltage changes dynamically, as opposed to the vast majority of datasets utilized in fuel cell modelling where the data comes either from static polarization curve experiments. Dynamical profiles of the operating conditions allow showing complex and hard to model behavior within the PEMFC.
2. The data-driven component of the hybrid structure is an ESN. As per the authors knowledge, this work represents the first implementation of an ESN for fuel cell modelling and voltage forecasting in a hybrid architecture.
3. A recursive GRBF is proposed for merging the contribution of each model. The GRBF recursively compares the similarity between the output of each model and the measured output of the fuel cell, assigns weights to each and computes the output of the hybrid structure.

The rest of this paper is organized as follows. Section **2** describes the experimental setup and explains the most important features of the dataset. Section **3** presents the simplified physics-based model. Section **4** presents the data-driven model, describe its fundamental parameters and the tuning procedure. Section **5** presents the hybrid structure and the merging approach. Section **6** shows and compares the results obtained from each model. Finally, conclusions and proposals for future research direction are presented in Section **7**.

## **2 Experimental setup and data**

The dataset used to build, validate and test the hybrid model comes from experimental tests conducted on an automotive PEM fuel cell stack as part of the INN-BALANCE European research project [20]. The objective of the experimental tests was to analyze the effect of dynamic changes in the operating conditions on the voltage profile of the stack. The PEMFC stack was fed with pure hydrogen and air. A total of 12 operating conditions could be controlled: inlet pressure in anode and cathode, outlet

pressure in anode and cathode, relative humidity of the inlet gases in anode and cathode, stoichiometry of hydrogen and air, inlet flow of hydrogen and air, stack temperature and load current. In each experiment one operating condition was changed dynamically and the rest were kept constant. The dataset used for the present work corresponds to experiments in which the cathode stoichiometry is swept downward in regular steps while all the other operating conditions are set approximately constant. The cell voltage have been normalized to the value of the theoretical voltage at standard conditions, that is 1.229 V; Fig. 1.

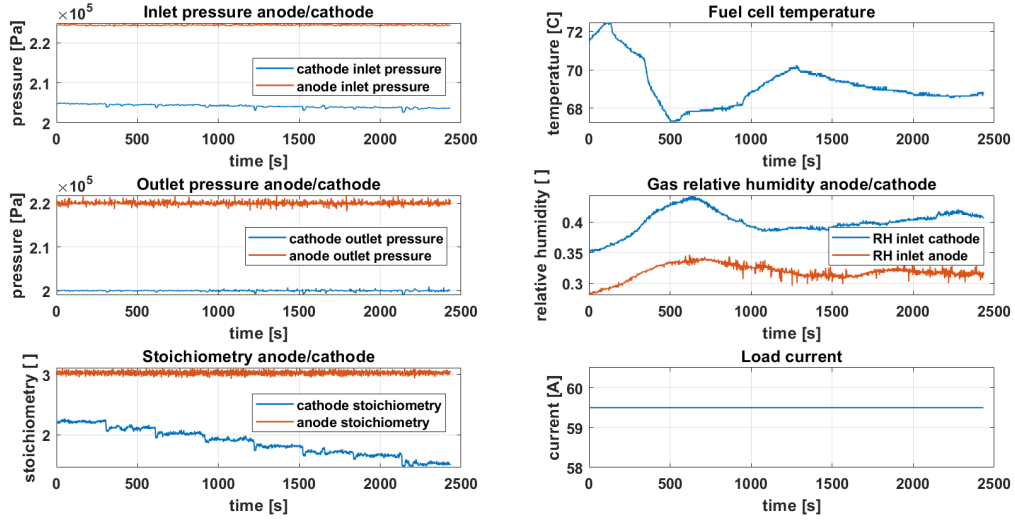


Figure 1: Operating conditions

A total of eight different conditions have been tested and are described in Tab.(1) shown below:

Anode stoichiometry set	Cathode stoichiometry set	Current density set	Condition number
[-]	[-]	[A]	[-]
3	2.2	200	0
3	2.1	200	1
3	2.0	200	2
3	1.9	200	3
3	1.8	200	4
3	1.7	200	5
3	1.65	200	6
3	1.5	200	7

Table 1: Table summarizing the several operating conditions tested

The resulting output voltage is shown in Fig. 2. The experiment shows highly nonlinear, hard to model behaviors of PEM fuel cells. As it has been shown by several studies, changing the cathode stoichiometry cause changes in the output voltage, in general, decreasing values of air stoichiometry result in decreasing values of voltage and vice versa. This respond has been studied mostly on static conditions through the analysis of polarization curves. However, as it can be seen in Fig. 2, during dynamic operation and within certain regions of current, the output voltage does not follow exactly the decreasing profile of the

stoichiometry. Instead, the voltage presents a recovery in some portions of the downward step.

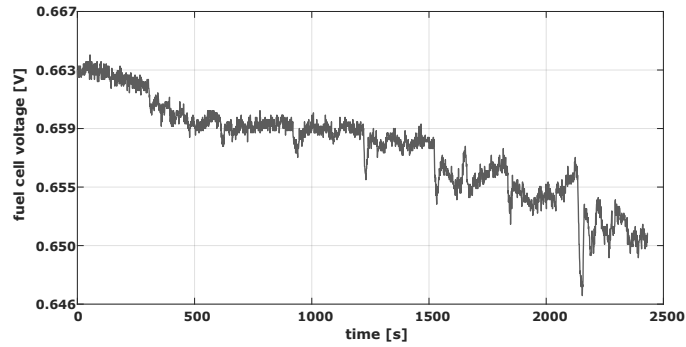


Figure 2: Fuel cell output voltage

The hypothetical explanation for this phenomenon is that depending on the value of other fundamental operating conditions such as cell temperature, relative humidity of the inlet gases and load current,

### 3 Physics-based model

#### 3.1 Electrochemical dynamics

The fuel cell voltage,  $v_c$ , can be approximated by the subtraction a series of voltage losses from the theoretical Nernst potential [21],

$$v_c = E_N - \frac{RT_c}{\alpha n F} \ln \left( \frac{i_c}{i_0} \right) - i_c R_{\text{ohm}} + \frac{RT_c}{n F} \ln \left( 1 - \frac{i_c}{i_{\text{lim}}} \right) \quad (1)$$

The cell current density  $i_c$  depends on the load and the cell temperature  $T_c$  is state and operating condition of the fuel cell that can be controlled through a cooling system.

Eq. (1) is composed by four clearly differentiated expressions, namely the Nernst voltage  $E_N$  and the voltage losses due to activation  $v_{\text{act}}$ , ohmic resistance  $v_{\text{ohm}}$  and reactant concentration  $v_{\text{con}}$ :

$$v_{\text{act}} = \frac{RT_c}{\alpha n F} \ln \left( \frac{i_c}{i_0} \right) \quad (2)$$

$$v_{\text{ohm}} = i_c R_{\text{ohm}} \quad (3)$$

$$v_{\text{con}} = \frac{RT_c}{n F} \ln \left( 1 - \frac{i_c}{i_{\text{lim}}} \right) \quad (4)$$

where the exchange current density  $i_0$ , the ohmic cell resistance  $R_{\text{ohm}}$ , and the limiting current  $i_{\text{lim}}$  are time varying parameters. In the case of the limiting current, it will be considered as a constant in the present work due to the region of operating current where the studied fuel cell is being operated. The

ideal gas constant  $R$ , the charge transfer coefficient  $\alpha$ , the number of attained electrons in the the oxygen reduction reaction  $n$ , and the Faraday constant  $F$ , are constants.

The Nernst voltage is a function of cell temperature and partial pressure of the reactant gases [22],

$$E_N = 1.229 - (8.5 \times 10^{-4})(T_c - T_{\text{ref}}) + (4.3085 \times 10^{-5})T_c(\ln P_{H_2} + 0.5 \ln P_{O_2}), \quad (5)$$

where  $T_{\text{ref}}$  is the reference temperature (298.15 K),  $P_{O_2}$  is the oxygen partial pressure in the cathode and  $P_{H_2}$  is the hydrogen partial pressure in the anode.

The exchange current density depends on operating conditions such as cell temperature and oxygen partial pressure:

$$i_0 = i_{0\text{ref}} A_c L_{Pt} \left( \frac{P_{O_2}}{P_{\text{ref}}} \right)^{0.5} e^{\left[ \frac{-\Delta G}{RT_c} \left( 1 - \frac{T_c}{T_{\text{ref}}} \right) \right]} \quad (6)$$

$L_{Pt}$  is the platinum loading in the catalyst layer in units of  $mg_{Pt}/cm^2$ ,  $A_c$  is the active catalyst area, also referred as the electrochemical surface area (ECSA) in  $cm^2/mg_{Pt}$ ,  $i_{0\text{ref}}$  is the reference exchange current density, a material specific constant parameter, and  $\Delta G$  is the Gibbs activation energy [23]. It has been widely proved that the exchange current density in the anode in a hydrogen fuel cell is three to four orders of magnitude larger than the exchange current density in the cathode, so the losses associated to activation energy in the anode are negligible [21]. Only cathode activation losses are taken into account in the simplified model.

The ohmic cell resistance is approximated to the membrane resistance in this work, and the ohmic losses of other regions of the fuel cell (i.e. bipolar plates) are neglected. Thus, the ohmic cell resistance (or ohmic membrane resistance) can be computed by Eqs. (7) to (8) per, [24]:

$$R_{\text{ohm}} = \frac{t_m}{\sigma_m} \quad (7)$$

where  $t_m$  is the constant membrane thickness, and  $\sigma_m$ , the membrane conductivity, is computed by:

$$\sigma_m = b_1 e^{\left( b_2 \left( \frac{1}{303} - \frac{1}{T_c} \right) \right)}$$

The coefficient  $b_2$  is constant and  $b_1$  is related to the water content of the membrane,  $\lambda_m$ , through:

$$b_1 = 0.005139 \lambda_m - 0.00326 \quad (8)$$

### 3.2 Gas channel dynamics

The oxygen partial partial pressure  $P_{O_2}$  is derived from the following mass balance analysis:

$$\frac{P_{O_2}}{P_{out}} = \frac{\dot{n}_{O_2}}{\dot{n}_{O_2} + \dot{n}_{N_2} + \dot{n}_v} \quad (9)$$

Being  $\dot{n}_{O_2}$ ,  $\dot{n}_{N_2}$  and  $\dot{n}_v$  the total molar flows of oxygen, nitrogen and vapor in the cathode,  $P_{out}$  is the total pressure of the air exiting the cathode, this value is provided in the dataset. The total flows of oxygen, nitrogen and water vapor are defined in Eqs. (10), (11) and (12) respectively:

$$\dot{n}_{O_2} = \dot{n}_{O_2,in} - \dot{n}_{O_2,ORR} \quad (10)$$

$$\dot{n}_v = \dot{n}_{v,in} + \dot{n}_{v,ORR} \quad (11)$$

$$\dot{n}_{N_2} = \dot{n}_{N_2,in} \quad (12)$$

where  $\dot{n}_{O_2,in}$ ,  $\dot{n}_{N_2,in}$  and  $\dot{n}_{v,in}$  are the inlet molar flows of oxygen, nitrogen and water vapor going into the cathode;  $\dot{n}_{O_2,ORR}$  and  $\dot{n}_{v,ORR}$  are the oxygen consumed and the water vapor produced by the oxygen reduction reaction (ORR) respectively. The inlet molar flow of vapor can be computed from the values of inlet air temperature and inlet air relative humidity provided in the dataset. First, the vapor saturation pressure and the vapor pressure of the air entering the cathode are computed through Eqs. (13) and (14) respectively,

$$P_{sat,air,in} = p_0 e^{\left(\frac{E_v}{RT_{air,in}}\right)} \quad (13)$$

$$P_{v,in} = RH_{air,in} P_{sat,air,in} \quad (14)$$

where  $P_{sat,air,in}$  is the vapor saturation pressure,  $P_{v,in}$  is the vapor pressure,  $T_{air,in}$  is the temperature,  $RH_{air,in}$  is the relative humidity of the inlet air and  $p_0$  is a fitting coefficient (Table 2). Then the partial pressure of the dry part of inlet air  $P_{dry,in}$  and molar mass of the inlet air  $M_{dry}$ ,  $M_x$  being the molar mass of x, are computed through Eqs. (15) and (16):

$$P_{dry,in} = P_{air,in} - P_{v,in} \quad (15)$$

$$M_{dry} = 0.21M_{O_2} + 0.79M_{N_2} \quad (16)$$

The humidity ratio, computed through Eq. (17), is then used to calculate the molar flow of vapor

entering the cathode  $\dot{n}_{v,\text{in}}$ , Eqs. (18) to (20):

$$w_{ca} = \frac{M_v}{M_{dry}} \frac{P_{v,\text{in}}}{P_{dry,\text{in}}} \quad (17)$$

$$\dot{m}_{dry,\text{in}} = \frac{1}{1 + w_{ca}} \dot{m}_{air,\text{in}} \quad (18)$$

$$\dot{m}_{v,\text{in}} = \dot{m}_{air,\text{in}} - \dot{m}_{dry,\text{in}} \quad (19)$$

$$\dot{n}_{v,\text{in}} = \frac{\dot{m}_{v,\text{in}}}{M_v} \quad (20)$$

The vapor molar flow produced in the ORR is computed through Eq. (21):

$$\dot{n}_{v,\text{ORR}} = \frac{N_c I_c}{2F} \quad (21)$$

The total oxygen molar flow is computed in a similar manner, Eqs. (22) to (24):

$$\dot{m}_{O_2,\text{in}} = x_{O_2,\text{in}} \dot{m}_{dry,\text{in}} \quad (22)$$

Where  $x_{O_2,\text{in}}$  is the mass fraction of oxygen in dry air. Then the molar flows of inlet oxygen and oxygen consumed in the ORR are:

$$\dot{n}_{O_2,\text{in}} = \frac{\dot{m}_{O_2,\text{in}}}{M_{O_2}} \quad (23)$$

$$\dot{n}_{O_2,\text{ORR}} = \frac{N_c I_c}{4F} \quad (24)$$

The values and units of constants used in the equations (9) to (24) are shown in Table 2. The approximation proposed for  $P_{O_2}$  may have, in the range of operation of the fuel cell under study, an estimation error of around 10% [25]. An analysis of the expression for the Nernst voltage, Eq. (5), shows that errors in the order 20% in the estimation of the oxygen partial pressure produce errors of around 1% in the computed value of the Nernst voltage [25].

### 3.3 Membrane dynamics

The membrane hydration model approximates the water content in the membrane. This value is used to compute the membrane ohmic resistance,  $R_{\text{ohm}}$ . The membrane average water content,  $\lambda_m$ , is calcu-



lated averaging the water activities in the anode  $a_{an}$  and cathode  $a_{ca}$  that can be calculated from the approximated vapor partial pressure, Eqs. 25 to 27, per [24]:

$$a_i = \frac{P_{v_i}}{P_{sat_i}} \quad (25)$$

for  $i = fan, cag$  and where  $P_{sat}$  or  $P_{sat,air,in}$  are calculated through Eq. (13).

$$a_m = \frac{a_{an} + a_{ca}}{2} \quad (26)$$

$$\lambda_m = 0.043 + 17.81a_m - 39.85a_m^2 + 36.0a_m^3 \quad (27)$$

The computed  $\lambda_m$  is then introduced in Eq. (8) to approximate the membrane ohmic resistance.

Symbol	Description	Value	Units
$p_0$	Fitting pressure coefficient	30.05	GPa
$E_v$	Vapor energy	36.98	kJ mol <sup>-1</sup>
$R$	Gas constant	8.314	J molK <sup>-1</sup>
$M_{O_2}$	Oxygen molar mass	32 · 10 <sup>-3</sup>	kg mol <sup>-1</sup>
$M_{N_2}$	Nitrogen molar mass	28 · 10 <sup>-3</sup>	kg mol <sup>-1</sup>
$M_v$	Water vapor molar mass	18 · 10 <sup>-3</sup>	kg mol <sup>-1</sup>
$i_{0ref}$	Reference exchange current density	2.9 · 10 <sup>-5</sup>	mA
$i_{lim}$	Limiting current	1760	mA
$\alpha$	Charge transfer coefficient	0.397	kJ mol <sup>-1</sup>
$n$	Number of electrons in the reaction	2	-

Table 2: Values and units of the constants

## 4 Data-driven model

In order to complete the information provided by the physics-based model, a data-driven model was used in parallel. The interest of this structure is to adjust the predictions made by the physical model by using the data-driven model to use historical data and capture more complex dynamics that are difficult to represent. The data-driven model is based exclusively on historical data. The computation is made by IA algorithm. Challenges are finding the best tool to catch the dynamics of the variable evolution with a low computational effort, the setting of the algorithm parameters and the quality of the data available for the training step. One of the best-known data-driven algorithms dedicated for forecasting task are the Recurrent Neural Networks (RNN). Indeed, they allow capturing temporal information at different scales which make them suitable for this forecasting task. However, classical optimization methods such as back-propagation do not provide good performances and can lead to vanishing gradient problems, where

the derivative calculated for some network weights becomes so small that no change is made between epochs. [26,27]. Another possibility is to use Long Short Term Memory (LSTM) [28] or Gated Recurrent Unit (GRU) [29] networks. The architecture of these networks helps to preserve the error that can be back-propagated through time and layers. LSTM cells are composed of three gates which are: input, output, and forget gates. They allow keeping or erasing the information stored in memory. The input gate represents new information added to the cell state, the forget gate decides what information is to be stored or deleted, while the output gate corresponds to the output of the LSTM. The principle of GRU cells is very similar to that of LSTMs. The input and forget gates are simplified to an update gate and a reset gate is used to control how much past information is to be forgotten. Since the architecture of GRU cells is simpler than that of LSTMs, they are more computationally efficient. According to the empirical study in [30], GRU cells are more attractive when the sequence is long and the data sets are small. In other scenarios, the performance losses compared to LSTMs are more serious. Fig. 3 resumes the principle of the 3 cells.

Despite improvements in RNNs architectures, training using back-propagation through time requires high computational resources. To compensate for this limitation, Reservoir Computing (RC) uses an approach that is significantly different. RC principle is to map one or more input signals into a high-dimensional computational space containing abundant dynamic transient states (called a reservoir). For this, the reservoir weights are fixed and only a readout is trained. It has been developed independently by Jaeger and Maas in the form of Echo State Networks (ESN) [31] and Liquid State Machines (LSM) [32] respectively in the early 2000s. In [33] a review of main reservoirs and readout training methods is proposed. Although the principle of the RC approaches is the same, differences can be found in the ways of initializing the reservoir and training the weights in the readout.

The data-driven model developed in this paper is based on the use of echo state networks. They have been widely used in several areas such as gesture recognition [34], motor control [35], sunspot activity prediction [36] and fuel cell remaining useful life prediction [37]. In addition, ESNs are easier to implement and tune than LSMs [33] which makes them interesting for embedded applications.

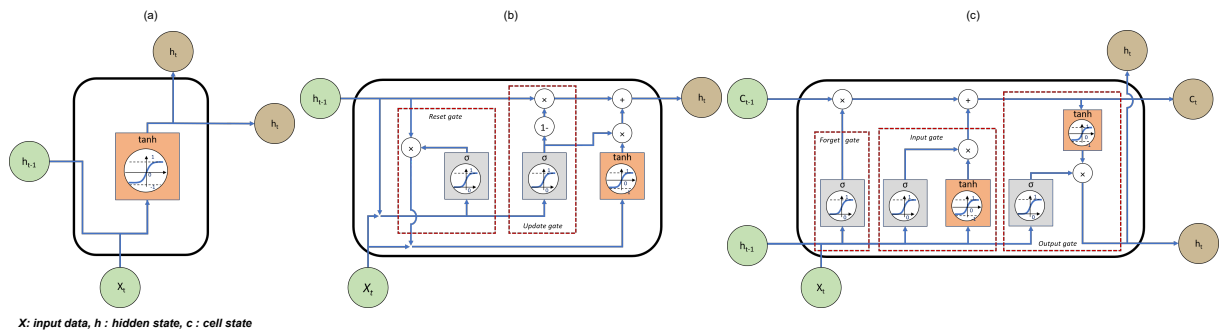


Figure 3: Presentation of main data-driven approaches, (a) RNN, (b) GRU, (c) LSTM

## 4.1 Echo State Networks

According to the work of Jaeger in [31], ESNs can be represented by three layers. First of all, the information is received by neurons in the input layer. Then, the information is propagated in a high dimensional space named reservoir composed of a high number of neurons which weights are fixed during the initialization. The last step is the transmission of the information from the reservoir to an output layer to read the results. Because the weights of input and recurrent connections are fixed during the generation of the reservoir, only the output weights are trained through linear regression. This feature reduces the complexity of the network and its computation time compared to the usual methods (i.e. LSTM and GRU). Therefore, ESN are able to process complex non linear task using a simple structure.

### 4.1.1 Equations

First of all, for a better understanding of the equations below, the terminology is fixed. In this study, a discrete-time Echo State Network with  $K$  input units (i.e features),  $N$  reservoir-internal units, and  $L$  output units considered. In addition, the discrete time is represented by  $n = 1, 2, \dots, T$  where  $T$  is the number of data points in the training database.

The input weights  $\mathbf{W}^{in}$  are collected in a matrix of size  $N \times K$ . The activation of the input neurons at time "n" are represented by the input vector:  $u(n) = (u_1(n), \dots, u_K(n))$ .

The reservoir weights  $\mathbf{W}^x$  are collected in a matrix of size  $N \times N$ . The activation of the reservoir neurons at time "n" are represented by the reservoir state vector:  $x(n) = (x_1(n), \dots, x_N(n))$ .

The output weights  $\mathbf{W}^{out}$  are collected in a matrix of size  $L \times (K + N + L)$ . The activation of the output neurons at time "n" is represented by the output vector:  $y(n) = (y_1(n), \dots, y_L(n))$ .

For specific applications an optional feedback weight matrix  $\mathbf{W}^{back}$  of size  $N \times L$  can be added between the output weight matrix and the reservoir.

The reservoir activation states are calculated using Eq. (28) described below:

$$x(n) = f(\mathbf{W}^{in} \cdot u(n) + \mathbf{W}^x \cdot x(n-1) + \mathbf{W}^{back} \cdot \mathbf{y}(n-1)) \quad (28)$$

where  $f$  and  $n$  represent respectively the activation function of the neurons and the time step. In tasks where no output feedback is required (i.e.  $\mathbf{W}^{back}$  is null), the activation of output neurons can be calculated using the result of Eq. (29) presented below:

$$y(n) = f_{out}(\mathbf{W}^{out} \cdot [u(n) \parallel x(n)]) \quad (29)$$

where  $f_{out}$  is the activation function for the output neurons (generally identity) and  $[u(n) \parallel x(n)]$  is the concatenation of  $u(n)$  with  $x(n)$ .

The output weights are calculated by solving the linear equation system described below:

$$\mathbf{W}^{out} \quad X = Y^{target} \quad (30)$$

where  $X$  represents all  $[u(n)jx(n)]$  produced by presenting the reservoir with  $u(n)$  and  $Y^{target}$  represents all  $y_{target}(n)$ .

When training the network, the objective is to minimize the Mean Squared Error (MSE) in Eq. (30) which is represented by Eq. (31) shown below:

$$MSE = \sum_{n=1}^T (y(n) - y_{target}(n))^2 \quad (31)$$

#### 4.1.2 Parameters

Besides their simplicity of use and optimization, ESNs come with a larger number of parameters to set which are: spectral radius, connectivity, leaky rate, amount of neurons, and scaling factor.

*Spectral radius:* The spectral radius is used to set the intensity at which past states affect current states. The higher the value, the more the past states will be reflected. Mathematically, it corresponds to the maximum eigenvalue of the reservoir matrix. It is determined during the initialization of the reservoir where the matrix weight is generated randomly. In order to let the user control this parameter, a general solution is to normalize the matrix by dividing the matrix by its spectral radius value (range scaling). Another way is to divide the matrix by its Euclidean norm (norm scaling). In both cases, the normalized matrices are then multiplied by the desired value. It is recommended to have a spectral radius value of less than 1 in order to respect the Echo State Property (ESP) which implies that the initial conditions should gradually disappear with time, i.e. the state of the reservoir should depend only on the input signal and not on the initial conditions existing before this input. According to the study done in [38], the ANOVA method has been applied to ESN and results show that spectral radius and the number of neurons are the most important parameters to define.

To take this limitation into account, in [39], authors proposed to use Glorot distribution (also named Xavier distribution) [40], the weights are directly initialized to follow a normal, truncated normal or uniform distribution depending on the number of inputs and outputs. This method eliminates the need to normalize weights using the spectral radius, thus reducing the number of parameters to be used. For normal and truncated normal distributions,  $W \sim N(\mu = 0, \sigma^2)$  the standard deviation ( $\sigma$ ) is calculated as shown in Eq. (32)

$$\sigma = \sqrt{\frac{2}{(n_{in} + n_{out})}} \quad (32)$$

For a uniform distribution  $W \sim U[bound, bound]$ , the bounds are determined as shown in Eq. (33):

$$bound = \sqrt{\frac{6}{(n_{in} + n_{out})}} \quad (33)$$

*Leaky rate:* The second parameter is the leaky rate  $\alpha$  which allows controlling the dynamics of neurons (also named leaky integrators neurons). The leaky value is in the interval  $[0, 1]$  and it represents the importance given to previous state of the reservoir to the calculation of current state. A high leakage rate signifies that previous state have a low impact on current outputs. To integrate the leakage, it is necessary to modify the calculation of neurons presented in Eq. (28) by Eq. (34) shown below:

$$x(n+1) = (1 - \alpha) \cdot x(n) + \alpha \cdot f(\mathbf{W}^{in} \cdot u(n) + \mathbf{W}^x \cdot x(n) + \mathbf{W}^{back} \cdot y(n)) \quad (34)$$

*Connectivity:* The third parameter is the connectivity ( $c$ ) which represents a percentage of non-zero weights in the reservoir matrix. Adding zeros within the matrix allows increasing individual dynamics by decoupling into sub-networks. According to Lukoševičius in [41], the impact of connectivity on the results is relatively small. However, a sparsely connected reservoir improves computation times due to the fact that reservoirs are updated faster.

*Neurons:* In opposition to classical recurrent neural networks (RNN, LSTM, GRU), ESN reservoirs have the capacity to process a large number of neurons. This number can vary from a dozen to several thousand. Indeed the weights being fixed allow simplifying computation times because the problem is transformed into a simple linear regression.

*Scaling factor* The scaling factor represents the interval in which the input and feedback weights of the network will be fixed during initialization. For a normal distribution, this value is characterized by its standard deviation, and for a uniform distribution by its interval  $[-a, a]$ . Typically, Echo State Networks are used with hyperbolic tangent activation function and scaling factors are in the interval  $[-1, 1]$  or a standard deviation of 0.5. Indeed, the closer the weights are to 0, the more linear the output will be because the hyperbolic tangent function is almost linear around 0.

### 4.1.3 Reservoir combination

In order to cope with the increase in the number of parameters to be regulated, some architectures combining several reservoirs have been developed. The first combination is the use of bidirectional reservoirs. In order to increase the number of detectable dynamics in a sequence, a solution is to double the number of reservoirs. This allows learning input sequences in both chronological and reverse directions and thus gives the same importance to the weights situated at the beginning and end of sequences. In [42], Bianchi used a neural network composed by a Bidirectional ESN (BiESN) and some feed-forward layers to classify time series. Bidirectionality can be used with all types of RNN approaches however, due to the computation times they are not often applied. The second combination is named "Multi-reservoir ESN" (MR ESN) and consists in using several small tanks instead of a large one. The tanks can be combined in series and/or in parallel, which makes it possible to use several configurations at the same time and

thus avoid having a phase dedicated to the optimization of the reservoir parameters. In [43], Sun presents the main existing architectures combining several ESN, and in [44], a Multi-Reservoir Bidirectional ESN (MR-BiESN) has been used to forecast the voltage of a PEMFC in a static operating condition.

## 4.2 Data-driven model designed

The main objective of the designed approach is to predict the future voltage of a PEMFC using past voltage and current. In order to reduce the computation time but also the training time (including the time needed to find the right combination of parameters), the use of a MR-BiESN seems to be a relevant choice. Indeed, as explained in the previous section, Setting the parameters values to obtain good results need both time and skills as it is done by a trial and error approach or by an optimization tool. Furthermore, in case of multiple dynamics to catch, the final set is a compromise to get the dominant one. An alternative is the use multiple tanks, in parallel or in series, instead of a single and large one. The parameters of the several sub tanks are sorted in the usual range of values : the neuron dynamics of the relevant subtank are amplified and the dynamics of the irrelevant ones are estinguished during the training of the ESN. In the designed model, three bidirectional reservoirs are used in parallel. The parallel organization allows detecting the different dynamics contained in the data. Each reservoir is set to detect a type of dynamics (low, medium or fast). The data-driven architecture is presented in Fig. 4 and selected parameters are described below and summarized in Table 3.

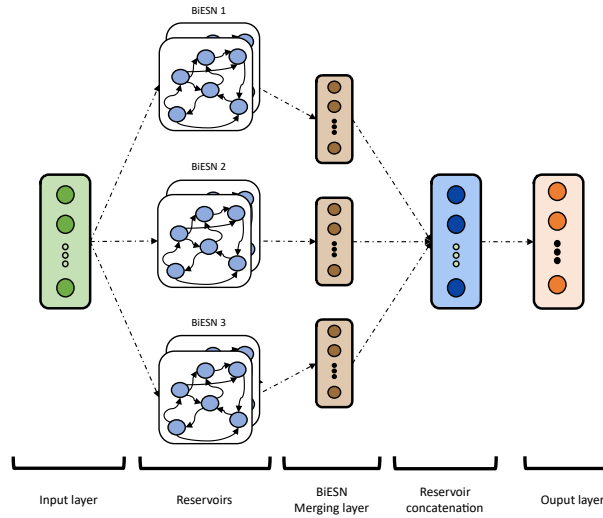


Figure 4: Architecture of the designed data-driven model

The dynamics of each reservoir are represented by a parametrization of the spectral radius and the leakage rate more or less large according to the desired dynamics. In this study, the retained values for both parameters are 0.1, 0.5, and 0.9 (extreme and average values). The connectivity remains constant at 10% which is a default value and the number of neurons is fixed at 100. The input and reservoirs' weights are initialized using a Glorot Uniform distribution. The advantage is that it directly takes into

account the number of inputs and outputs and avoids having to search for a good value to bound the uniform distribution.

Inputs of the network are sequences representing the past voltage and cathode stoichiometry set by the user. An important parameter to define is the length of the sequence to be used, indeed, a too-short sequence cannot use the information of the past and a too-long sequence increases the complexity and the computing time. In this study, it has been chosen to train the network using the two last minutes with a sampling of 1Hz. An important assumption is that in order to predict the voltage at time  $t+1$ , it is assumed that the user-defined cathode stoichiometry at time  $t+1$  is known. In a system, depending of the sampling frequency, this value can be estimated close to the last known value. This makes that there is a shift in the learning sequences as shown in Fig. 5.

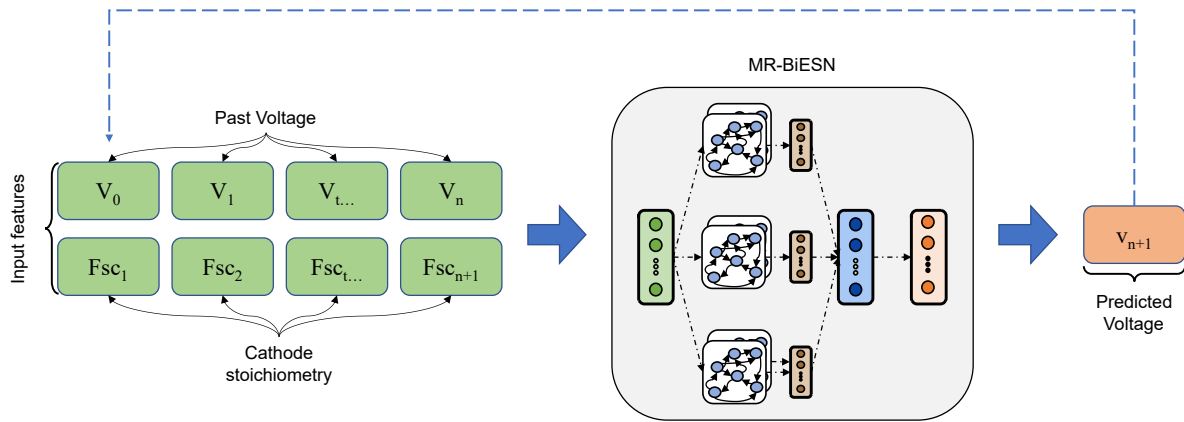


Figure 5: Scheme representing the functioning of the data-driven model

To improve performances and reduce optimization complexity, each feature has been normalized in the range  $[0, 1]$ . The activation function used in all reservoirs is the hyperbolic tangent ( $\tanh$ ) which is the default function for ESN. The selected optimization algorithm is the one presented in [45] entitled AdamW. The authors demonstrated that it is more interesting to decouple the weight decay parameter with the Adam optimizer in order to facilitate the search for parameters but also to obtain better results. The weight decay is a regularization hyper-parameter used to improve the generalization of models by avoiding over-fitting. The selected learning rate and weight decay are respectively 0.0012 and 0.001. During the training, the two parameters have been scheduled to follow an exponential decay with a decay rate of 0.95 for each 10 epochs. It has been shown empirically that learning rate decay improves generalization and helps the optimization to reach a global optimum. Since the weight decay is decoupled, it is necessary to apply the same scheduling.

Parameters	Values	Comments
Spectral radius ( $\rho$ )	Not defined	-
Leaky rate ( $\alpha$ )	0.1 / 0.5 / 0.9	Low / Medium / High dynamic
Connectivity ( $c$ )	0.1	-
Number of neurons	For each reservoir: 100 Total: 600	-
Reservoirs initialization	Glorot Uniform	-
Reservoir weight scaling method	No scaling	-
Readout layer initialization	Glorot Uniform	-
Merging mode of bidirectional reservoirs	Concatenation	-
Activation function	tanh	-
Optimization	AdamW	-
Learning rate	0.0014	-
Weight decay	0.001	-
Scheduler (learning rate + weight decay)	Exponential decay decay rate : 0.95 decay steps: 10 epochs	-
Batch size	32	-
Sequences length	120	Last 2 minutes (120 seconds)

Table 3: Table summarizing the parameters retained to design the data-driven model

## 5 Hybrid model

The hybrid model is formed by three main components, the physic-based model, the data-driven model and the merging function, Fig. 6.

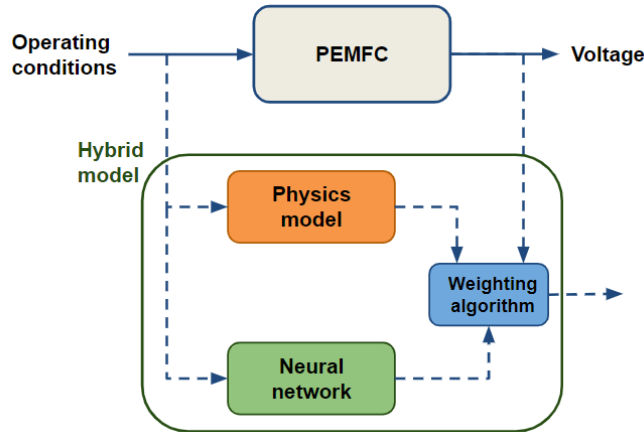


Figure 6: Proposed hybrid model formed by a simplified physics-based model and a neural network.

### 5.1 Merging function

The merge of models is based on a similarity function, the Gaussian radial basis function. This function measures the distance between measured (real) fuel cell voltage and the voltage approximated by the model, and assigns a weight,  $w_{model}$ , in the range  $[0, 1]$ , to the voltage approximation at each sampling time:



$$w_{model}(l) = \exp\left(-\frac{[v_{model}(l) - v_{real}(l)]^2}{2\lambda^2}\right) \quad (35)$$

for

$$l = \{k - N_P, k - (N_P - 1), k - (N_P - 2), \dots, k - (N_P - N_P)\}$$

$N_p$  is the number of past samples that are taken into account to compute the weight of the approximation,  $\lambda$  is a tuning parameter related to the confidence of the model,  $k$  is the current sampling time,  $v_{model}$  is the output voltage of the model (either data-driven or physics-based) and  $v_{real}$  is the measured fuel cell voltage.

Then, the  $N_P$  weights of each model are averaged for computing the contribution of each model to the prediction:

$$v_{hybrid}(n) = \bar{w}_{dd}v_{dd}(n) + \bar{w}_{pb}v_{pb}(n) \quad (36)$$

for

$$n = \{k + 1, k + 2, k + 3, \dots, k + N_F\}$$

Where  $\bar{w}$  is the averaged weight of each model, either data-driven (dd) or physics-based (pb),  $v$  is the voltage approximation of each model, and  $N_F$  is the prediction horizon.

## 6 Results and discussion

### 6.1 Physics-based model

Fig. 7 presents the results of the physics-based model. It can be seen that even though the difference between real and modeled voltage is in the range of fewer than ten millivolts, the model fails in following the trends and short-term dynamics of the voltage signal.

The difference in the trends at some sampling times can be explained by the simplifications and approximations done in the construction of the physics-based model, these include: overlooking the effect of several water dynamics that take place in the cathode catalyst layer, the approximations done in the equations that calculate the partial pressure of oxygen and hydrogen in the gas channels, the assumption that the fitting parameters are time-invariant and, in general, the fact that the model has been fitted to quasi-static operating point in the PEMFC (the polarization curve). The mean square error (MSE) of the model is shown in Table 4.

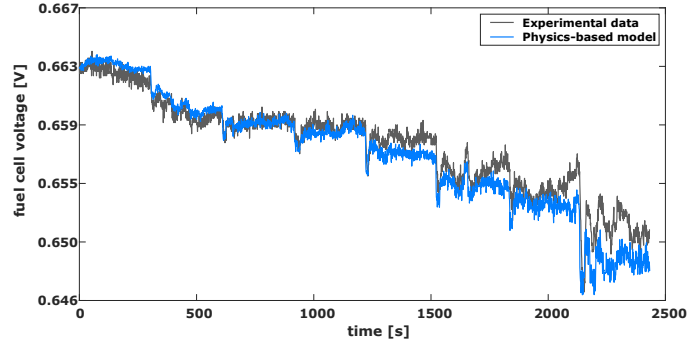


Figure 7: Real voltage and physics-based model predicted voltage.

## 6.2 Data-driven model

Regarding the data-driven model, it has been trained using the data presented in Section 2. The objective of the proposed approach is to predict the next 60 seconds from the last 120 seconds. For that purpose, a new signal is constructed from the training data in which the order of the tested conditions has been changed. The interest of reconstructing the signal is that it allows to measure the performance of the network during the transitions between two conditions. Moreover, to ensure that the model is able to generalize to intermediate conditions, conditions 2 and 5 presented in table 3 have been removed from the training base and will be used to test the model. This new reconstructed signal and the prediction results are presented in Fig. 8.

The results show that the data-driven method is able to correctly capture the voltage evolution trends while suppressing the noise observable during the signal measurement. The total MSE obtained is analyzed in table 4. Furthermore, the generalization condition was met as the model correctly predicts conditions 2 and 5 (600-900 sec and 1800-2100 sec) that were not learned.

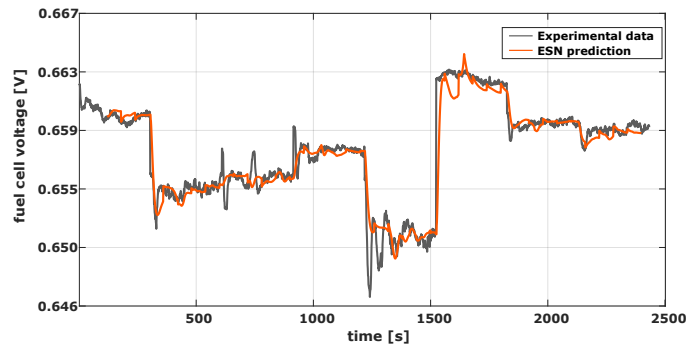


Figure 8: Real voltage and physics-based model predicted voltage. Conditions presented in Table 1 have been shuffle in the order: [1, 6, 5, 4, 7, 0, 2, 3]

### 6.3 Hybrid model

A new voltage profile is constructed by concatenating the experimentally measured voltage and the shuffled voltage signal that was used for testing the data-driven model. This concatenated signal represents a full cycle for the hybrid model, Fig. 9. In the first phase (from  $t = 0\text{s}$  to  $t = 2555\text{s}$ ), during which the ESN is being trained, the data-driven model does not provide any valid voltage forecast and the voltage approximation is done exclusively by the physics-based model. Then, in the second phase, once the neural network has been trained, its forecast can be incorporated into the hybrid model and combined with the physics-based model. The discrimination and weighting of both models is done by the merging algorithm.

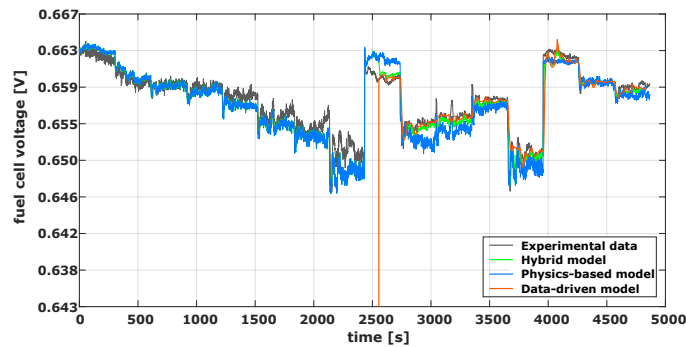


Figure 9: Cell voltage and models voltage approximation

It is expected that the two phases shown in Fig. 9, which correspond to training and prediction period of the ESN occur periodically during the fuel cell lifetime. This is due to the fact that, when facing completely new operating conditions or when the internal parameters of the fuel cell suffer large variations due to degradation (i.e. aging), the ESN would need to be re-trained. This highlights an important advantage of the proposed hybridization, that, by incorporating independent models it is able to maintain a reasonable approximation of the voltage during the full cycle of the fuel cell.

The modeling error of the three models is shown in Table 4. The MSE during each phase and the total MSE is calculated. It can be seen that, during the prediction phase, when both models contribute to approximate the output of the real fuel cell, the proposed hybridization achieves a lower modeling MSE than each of the models separately. The total MSE takes into account the whole cycle (training and prediction).

Model	Training	Prediction
Physics-based	$2.6247 \cdot 10^6$	$2.9339 \cdot 10^6$
Data-driven	-	$1.7777 \cdot 10^6$
Hybrid	$2.6247 \cdot 10^6$	$9.5692 \cdot 10^7$

Table 4: Mean square error of the different models at each phase of the cycle

## 7 Conclusions

Fuel cells are promising energy conversion devices in a future hydrogen-based economy. However, issues such as efficiency and durability must be addressed and the way of addressing them is by optimal control of fuel cell operating conditions. The performance of such control depends largely on efficient and accurate models. This paper has proposed and tested a novel hybrid, data-driven/physics-based model for approximating fuel cell output voltage. The two models are merged through a recursive Gaussian radial basis function. The hybrid model has been trained, validated, and tested with an experimental dataset, with dynamical profiles in the operating conditions. Future research will focus on the implementation of prediction models within control algorithms.

## Acknowledgements

This work has received support by the following sources: the Project MDM-2016-0656 and the program "ESF Investing in your future" [grant number BES-2017-082441] funded by MCIN/AEI /10.13039/501100011033. The European Union through project INN-BALANCE (grant number 735969) and project RUBY (grant number 875047, website: <https://www.rubyproject.eu>), both of the Fuel Cells and Hydrogen 2 Joint Undertaking. The contract "GEPEVE: Gestión predictiva de energía para eficiencia energética en vehículos eléctricos con control de crucero adaptativo" of the Movilidad 2030 project of the 2020 Misiones CDTI program. The Consolidated Research Group RAIG (2021 SGR 00510) of the Departament de Recerca i Universitats de la Generalitat de Catalunya. The doctoral school EIPHI (contract ANR-17- EURE-0002). And by the Region Bourgogne Franche-Comté.

## Conflict of interests

The authors declare that there is no conflict of interest and no financial competing interests.

## References

- [1] U.S. Department of Energy, Fuel Cell Technologies Office, "Multi-year research, development, and demonstration plan," 2014. [www.energy.gov/eere/fuelcells/downloads/hydrogen-and-fuel-cell-technologies-office-multi-year-research-development](http://www.energy.gov/eere/fuelcells/downloads/hydrogen-and-fuel-cell-technologies-office-multi-year-research-development).
- [2] R. Darling and J. Meyers, "Kinetic model of platinum dissolution in PEMFCs," *Journal of the Electrochemical Society*, vol. 150, no. 11, pp. A1523–A1527, 2003.
- [3] A. Pandey, Z. Yang, M. Gummall, V. Atrazhev, N. Y. Kuzminyh, V. Sultanov, and S. Burlatskyb, "A carbon corrosion model to evaluate the effect of steady state and transient operation of a polymer

- electrolyte membrane fuel cell,” *Journal of the Electrochemical Society*, vol. 160, no. 9, pp. F972–F979, 2013.
- [4] P. Schneider, C. Sadeler, A. Scherzer, N. Zamel, and D. Gerteisen, “Fast and reliable state-of-health model of a PEM cathode catalyst layer,” *Journal of The Electrochemical Society*, vol. 166, no. 4, pp. F322–F333, 2019.
- [5] A. Kneer and N. Wagner, “A semi-empirical catalyst degradation model based on voltage cycling under automotive operating conditions in PEM fuel cells,” *Journal of The Electrochemical Society*, vol. 166, no. 2, pp. F120–F127, 2019.
- [6] E. Lechartier, E. Laffly, M.-C. Péra, R. Gouriveau, D. Hissel, and N. Zerhouni, “Proton exchange membrane fuel cell behavioral model suitable for prognostics,” *International Journal of Hydrogen Energy*, vol. 40, pp. 8384–8397, 2015.
- [7] Z. Denga, Q. Chena, L. Zhanga, Y. Zongb, K. Zhoua, and Z. Fua, “Control oriented data driven linear parameter varying model for proton exchange membrane fuel cell systems,” *Applied Energy*, vol. 277, p. 115540, 2020.
- [8] A. Mohammadi, G. Cirrincione, A. Djerdir, and D. Khaburi, “A novel approach for modeling the internal behavior of a pemfc by using electrical circuits,” *International Journal of Hydrogen Energy*, vol. 43, pp. 11539–11549, 2018.
- [9] P. H. A. Nóbrega, “A review of physics-based low-temperature proton-exchange membrane fuel cell models for system-level water and thermal management studies,” *Journal of Power Sources*, vol. 558, p. 232585, 2023.
- [10] J. Wang, Y. Li, R. X. Gao, and F. Zhang, “Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability,” *Journal of Manufacturing Systems*, vol. 63, pp. 381–391, 2022.
- [11] Y. Meng, S. Yu, J. Zhang, J. Qin, Z. Dong, G. Lu, and H. Pang, “Hybrid modeling based on mechanistic and data-driven approaches for cane sugar crystallization,” *Journal of Food Engineering*, vol. 257, pp. 44–55, 2019.
- [12] M. von Stosch, R. Oliveira, J. Peres, and S. F. de Azevedo, “Hybrid semi-parametric modeling in process systems engineering: Past, present and future,” *Computers and Chemical Engineering*, vol. 60, pp. 86–101, 2014.
- [13] M. Hotvedt, B. Grimstad, and L. Imsland, “Developing a hybrid data-driven, mechanistic virtual flow meter - a case study,” *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 11692–11697, 2020.

- [14] M. Liang and S. Li, “Enhancing bnn structure learning of hybrid modeling strategy for free formulated mechanism complex systems,” *Journal of Process Control*, vol. 120, pp. 44–67, 2022.
- [15] N. Gaw, A. Hawkins-Daarud, L. S. . Hu, H. Yoon, L. Wang, Y. Xu, P. R. Jackson, K. W. Singleton, L. C. Baxter, J. Eschbacher, A. Gonzales, A. Nespodzany, K. Smith, P. Nakaji, J. R. Mitchell, K. R. S. Teresa Wu1, , and J. Li, “Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric mri,” *Sci Rep*, vol. 9, p. 10063, 2019.
- [16] F. Hua†, S. Hautaniemi, R. Yokoo, and D. A. Lauffenburger, “Integrated mechanistic and data-driven modelling for multivariate analysis of signalling pathways,” *Journal of the Royal Society Interface*, vol. 3, p. 515–526, 2006.
- [17] R. E. Baker, J.-M. Peña, J. Jayamohan, and A. Jérusalem, “Mechanistic models versus machine learning, a fight worth fighting for the biological community?,” *Biology Letters*, vol. 14, p. 0660, 2017.
- [18] L. Liao and F. Köttig, “A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction,” *Applied Soft Computing*, vol. 44, pp. 191–199, 2016.
- [19] G. Revach, N. Shlezinger, R. J. G. van Sloun, and Y. C. Eldar, “Kalmannet: Data-driven kalman filtering,” in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3905–3909, 2021.
- [20] INN-BALANCE Project Consortium. [www.innbalance-fch-project.eu](http://www.innbalance-fch-project.eu).
- [21] A. Dicks and D. Rand, *Fuel Cell Systems Explained*. Hoboken, NJ, USA: Wiley, 3rd ed., 2018.
- [22] J. T. Pukrushpan, H. Peng, and A. G. Stefanopoulou, “Control-Oriented Modeling and Analysis for Automotive Fuel Cell Systems ,” *Journal of Dynamic Systems, Measurement, and Control*, vol. 126, no. 1, pp. 14–25, 2004.
- [23] F. Barbir, *PEM Fuel Cells: Theory and Practice*. Elsevier Academic Press, 2005.
- [24] A. G. S. Jay T. Pukrushpan and H. Peng, *Control of Fuel Cell Power Systems*. London, U.K.: Springer, London, 1 ed., 2004.
- [25] J. Luna, E. Usai, A. Husar, and M. Serra, “Nonlinear observation in fuel cell systems: A comparison between disturbance estimation and high-order sliding-mode techniques,” *International Journal of Hydrogen Energy*, vol. 41, no. 43, pp. 19737–19748, 2016.
- [26] T. Kim and B. R. King, “Time series prediction using deep echo state networks,” vol. 32, no. 23, pp. 17769–17787.

- [27] H. Jaeger, “The” echo state” approach to analysing and training recurrent neural networks-with an erratum note’,” vol. 148.
- [28] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” vol. 9, no. 8, pp. 1735–1780.
- [29] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches.”
- [30] S. Yang, X. Yu, and Y. Zhou, “LSTM and GRU neural network performance comparison study: Taking yelp review dataset as an example,” in *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, pp. 98–101.
- [31] H. Jaeger, “The” echo state” approach to analysing and training recurrent neural networks-with an erratum note’,” vol. 148.
- [32] W. Maass, T. Natschläger, and H. Markram, “Real-time computing without stable states: A new framework for neural computation based on perturbations,” vol. 14, no. 11, pp. 2531–2560.
- [33] M. Lukoševičius and H. Jaeger, “Reservoir computing approaches to recurrent neural network training,” vol. 3, no. 3, pp. 127–149.
- [34] D. Jirak, S. Tietz, H. Ali, and S. Wermter, “Echo state networks and long short-term memory for continuous gesture recognition: a comparative study,”
- [35] M. Salmen and P. Ploger, “Echo state networks used for motor control,” in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pp. 1953–1958. ISSN: 1050-4729.
- [36] F. Schwenker and A. Labib, “Echo state networks and neural network ensembles to predict sunspots activity,” p. 6.
- [37] R. Mezzi, S. Morando, N. Y. Steiner, M. C. Péra, D. Hissel, and L. Larger, “Multi-reservoir echo state network for proton exchange membrane fuel cell remaining useful life prediction,” in *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, pp. 1872–1877. ISSN: 2577-1647.
- [38] S. Morando, S. Jemei, D. Hissel, R. Gouriveau, and N. Zerhouni, “ANOVA method applied to proton exchange membrane fuel cell ageing forecasting using an echo state network,” vol. 131, pp. 283–294.
- [39] Z. Carmichael, H. Syed, S. Burtner, and D. Kudithipudi, “Mod-DeepESN: Modular deep echo state network,” in *2018 Conference on Cognitive Computational Neuroscience, Cognitive Computational Neuroscience*.

- [40] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Y. W. Teh and M. Titterton, eds.), vol. 9 of *Proceedings of Machine Learning Research*, (Chia Laguna Resort, Sardinia, Italy), pp. 249–256, PMLR, 13–15 May 2010.
- [41] M. Lukoševičius, “A practical guide to applying echo state networks,” in *Neural Networks: Tricks of the Trade* (G. Montavon, G. B. Orr, and K.-R. Müller, eds.), vol. 7700, pp. 659–686, Springer Berlin Heidelberg. Series Title: Lecture Notes in Computer Science.
- [42] F. M. Bianchi, S. Scardapane, S. Løkse, and R. Jenssen, “Bidirectional deep-readout echo state networks,”
- [43] C. Sun, M. Song, S. Hong, and H. Li, “A review of designs and applications of echo state networks,”
- [44] D. Chanal, N. Y. Steiner, D. Chamagne, and M.-C. Pera, “Voltage prognosis of PEMFC estimated using multi-reservoir bidirectional echo state network,” in *2022 10th International Conference on Systems and Control (ICSC)*, pp. 352–359. ISSN: 2379-0067.
- [45] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,”