



HAL
open science

Tight hybridization of long short-term memory neural networks and robust Kalman filter for remaining useful life estimation of proton exchange membrane fuel cells

Mathias Lecroart, Audrey Giremus, Tudor-Bogdan Airimitoiaie, Mathieu Chevrié, Christophe Farges

► To cite this version:

Mathias Lecroart, Audrey Giremus, Tudor-Bogdan Airimitoiaie, Mathieu Chevrié, Christophe Farges. Tight hybridization of long short-term memory neural networks and robust Kalman filter for remaining useful life estimation of proton exchange membrane fuel cells. 2024. hal-04625867

HAL Id: hal-04625867

<https://hal.science/hal-04625867v1>

Preprint submitted on 26 Jun 2024

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.

Tight hybridization of long short-term memory neural networks and robust Kalman filter for remaining useful life estimation of proton exchange membrane fuel cells

Mathias Lecroart

*Univ. Bordeaux, CNRS, Bordeaux INP
IMS, UMR 5218*

Talence, France

mathias.lecroart@ims-bordeaux.fr

Audrey Giremus

*Univ. Bordeaux, CNRS, Bordeaux INP
IMS, UMR 5218*

Talence, France

audrey.giremus@ims-bordeaux.fr

Tudor-Bogdan Airimitoiaie

*Univ. Bordeaux, CNRS, Bordeaux INP
IMS, UMR 5218*

Talence, France

tudor-bogdan.airimitoiaie@ims-bordeaux.fr

Mathieu Chevrié

*Univ. Bordeaux, CNRS, Bordeaux INP
IMS, UMR 5218*

Talence, France

mathieu.chevrie@ims-bordeaux.fr

Christophe Farges

*Univ. Bordeaux, CNRS, Bordeaux INP
IMS, UMR 5218*

Talence, France

christophe.farges@ims-bordeaux.fr

Abstract—As part of industry 4.0, predictive maintenance aims to achieve the best compromise between risk prevention and repair costs. It leverages diagnostic and prognostic methods that can be either model-based or data-driven. More recently, a combination of both has been considered. This paper deals with the estimation of the remaining useful life of a proton exchange membrane fuel cell by predicting the evolution of its power over time. We propose to combine a Kalman Filter (KF) with a Long Short-Term Memory (LSTM) neural network. Contrary to recent works, a tight coupling is considered. The LSTM neural network is trained to learn only a part of the KF prediction model whereas the remaining dynamics are described by knowledge models. Moreover, the available data exhibit outliers due to interruptions during the measurement campaign. To mitigate their impact, a robust version of the KF and a robust cost function for the training of the LSTM are considered as a second contribution.

Index Terms—Hybrid models, Kalman Filter, long short-term memory neural network, Remaining Useful Life estimation, fuel cells, predictive maintenance.

I. INTRODUCTION

Predictive maintenance is a concept that has been known for many years but has become only recently feasible from an economic and technological perspective. It is a compromise between preventive and corrective maintenance with the objective of saving costs while ensuring safety [1]. In this paper, we focus on estimating the remaining useful life (RUL) of proton exchange membrane fuel cells (PEMFC). The latter represent an alternative to fossil fuels as they generate electricity from hydrogen and do not emit carbon dioxide. However, their limited lifespan hinders significant industrial

deployment. Prognostics of their RUL offers a solution to alleviate this difficulty by enabling preventive actions that extend their operational duration. It can be achieved by predicting the PEMFC power degradation. The RUL is then defined as the elapsed time until it reaches a critical threshold.

Many prediction algorithms exist. They can be based on knowledge models that take the form of state space representations, such as classically Kalman filters (KF) [2] or particle filters [3]. Alternatively, deep learning uses past data to train an evolution model. Long short-term memory (LSTM) networks [4] are widely used to manage sequential data [5] [6] and transformers have yielded promising results in some fields such as large language models [7]. However, on the one hand, for knowledge models, it is challenging to strike a balance between generalization and high estimation accuracy. On the other hand, data-driven models require rich databases, which can be difficult to obtain depending on the application. More recently, it has been proposed to take advantage of both types of approaches by coupling them. Hybridization makes it possible to estimate completely or partially unknown behaviors in the evolution model while remaining attached to expert knowledge.

Hybrid architectures have been considered in a variety of applications. Most of them involve relatively simple couplings where deep learning either replaces the dynamic or the observation equation of a state space representation, or provides pre-processed measurements. For instance, in order to predict hurricanes, Qin et al. [8] propose cascading an LSTM with a KF, with the output of the former being used as a measure in the KF correction step. Reference [9] employs a similar hybridization between a transformer network and an unscented

particle filter [10]. Also, in [11], [12] and [13], it is suggested replacing the state prediction step of a KF with an LSTM. This method enables good short-term estimations.

The use of hybrid models for PEMFC RUL estimation is widespread for two reasons: there is no evolution model that takes into account all the degradations of the cell, and it is complicated and costly to obtain extensive degradation data. In [14], it is proposed to filter voltage degradation measurements of a PEMFC using an Adaptive Extended KF (AEKF) while estimating residual voltage using a Nonlinear AutoRegressive eXogenous Neural Network (NARXNN). Then, the outputs of both algorithms are summed to predict future voltage. Similarly, Xia et al. [15] use a regression algorithm to separate the measurement into a low-frequency behavior predicted by an AEKF and a residual, high-frequency behavior predicted by an LSTM. Once again, the combination is made *a posteriori*. Deep learning models can also be leveraged during the correction step of a filter, as illustrated in [16], wherein the output of an LSTM-attention network serves as an online Kalman observation to correct the state-of-charge prediction of a PEMFC. Finally, a popular hybrid architecture is to predict the prior state thanks to an LSTM network such as in [17] or [6]. If the degradation behavior is fully unknown, this method is interesting to bypass expert knowledge.

This literature review prompted us to ponder whether a more tightly focused and less generalized hybridization could be achieved. In this paper, we propose a tighter coupling approach wherein an LSTM is integrated at the prediction step of a Kalman filter but only handles a part of the state variables. More precisely, it propagates the latter while estimating their cross-correlation with the remaining variables which evolve according to classical dynamic equations. The choice of an LSTM network rather than a transformer is motivated by the need to process our data sequentially, which the transformer does not naturally do. Indeed, a positional index should be added to the data, which would require a pre-processing step and heavier operations. This increased cost is not necessary in our context wherein the hybridization already alleviates the difficulty of the learning task. Finally, PEMFC databases include outliers that are due to the ageing process implemented during the measurement campaigns. To accommodate them, a solution could be to inflate the measurement variance at the cost of a loss of estimation accuracy. To avoid such a compromise, the proposed approach is made robust in two complementary ways: a Huber lost function is used for the LSTM training while a M-estimator version of the KF [18] is considered.

This article is organized in four sections. After this introduction, section 2 is dedicated to the considered hybrid algorithm to estimate PEMFC power over time. Section 3 illustrates the performance of the latter by testing it on a reference PEMFC database. It also includes comparisons with alternative architectures. Finally, section 4 presents conclusions and perspectives.

II. PROPOSED RUL ESTIMATION METHOD

A. PEMFC power modelling

PEMFC degradation monitoring has been studied in different papers. A comprehensive review of existing techniques and an introduction to PEMFC are proposed in [19] where it is shown that the power at time step k can be expressed as a function of the PEMFC physical characteristics as follows:

$$P_k = nI_k \left[E_{\text{rev}} - \frac{C}{\alpha_a} \ln \left(\frac{i_{\text{loss}} e^{b_{\text{loss}} k T_e} + \Gamma_k}{i_{0,a}} \right) - \frac{C}{2\alpha_c} \ln \left(\frac{i_{\text{loss}} e^{b_{\text{loss}} k T_e} + \Gamma_k}{i_{0,c}} \right) - \Gamma_k (R_{\text{ion}} e^{b_{\text{ion}} k T_e} + R_k) + B_k \ln \left(1 - \frac{\Gamma_k}{\frac{2}{C} D_k} \right) \right] - p \quad (1)$$

where $C = \frac{R_a T}{2F_a}$ with R_a the universal gas constant, F_a the Faraday constant and T is the stack temperature. E_{rev} is the nominal voltage and T_e is the sampling period. α_a and α_c are the charge transfer coefficients at the electrodes, i_{loss} represents the internal currents within the stack whereas $i_{0,a}$ and $i_{0,c}$ are the exchange current densities at each electrode, R_{ion} is the ionic resistance and b_{ion} , b_{loss} are degradation coefficients while p is a stack correction. In (1), the parameters indexed with k are time-varying while the others are constant. The latter can be identified using for instance a particle swarm optimization algorithm [19]. As for the dynamic parameters B_k , D_k and R_k , which respectively represent two diffusion and a contact resistance degradation, their evolution can be well described by a uniform rectilinear motion model. Conversely, the behaviour of the active area deterioration over time, represented by Γ_k , is not well-known whereas it plays a key role on the power variation.

The estimation of the RUL can be performed by inferring the time-varying parameters of (1) from a series of power measurements. To this end, a non-linear recursive filter such as an extended KF (EKF) or an Iterative KF (IKF) is well-suited [19]. After this learning step, the filter can be run as a predictor only to extrapolate the PEMFC power for future time instants. The RUL is then obtained as the elapsed time until the latter becomes lower than a threshold. The KF and its variants are interesting prognostic tools in the sense that they provide a measure of uncertainty associated to the estimates. However, they require an accurate state model of the unknown parameters and may not perform optimally with non-Gaussian noise distributions. Both issues are addressed in the sequel.

B. State space representation

Recursive filters are based on a state space representation of the parameters of interest, which are stacked in a state vector denoted \mathbf{x}_k in the sequel. It comprises both a prior evolution model of the latter, called the state equation, and an observation equation that relates it to a series of measurements referred to as $\{y_k\}_{k \geq 1}$. In the considered setting, the state vector can be decomposed in two parts: $\mathbf{x}_k^1 = (B_k, \dot{B}_k, D_k, \dot{D}_k, R_k, \dot{R}_k)^\top$ gathers the parameters with a known dynamic behaviour,

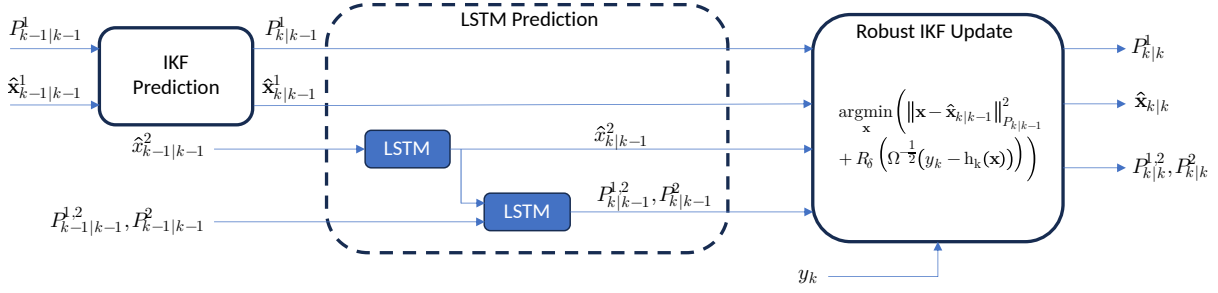


Fig. 1. Tight integration of LSTM units in a robust iterative KF (IKF)

whereas $x_k^2 = \Gamma_k$ corresponds to the parameter whose evolution is assumed completely unknown. It should be noted that the derivatives of B_k , D_k and R_k have to be appended to the state vector in the case of uniform rectilinear motions. The state variables satisfy the following discrete-time recursive propagation model:

$$\mathbf{x}_k^1 = F_k^1 \mathbf{x}_{k-1}^1 + \mathbf{v}_k^1 \quad (2)$$

$$x_k^2 = f_k^2(x_{1:k-1}^2, v_k^2) \quad (3)$$

where $x_{1:k-1}^2 = (x_1^2, \dots, x_{k-1}^2)$ and $\mathbf{x}_k = \left((\mathbf{x}_k^1)^\top, x_k^2 \right)^\top$. The random processes $\{\mathbf{v}_k^1\}_{k \geq 0}$ and $\{v_k^2\}_{k \geq 0}$ are mutually independent white noises. Under the assumption that the diffusion and contact resistances are independent, the transition matrix F_k^1 and the covariance matrix of \mathbf{v}_k^1 , denoted Q_k^1 , are block-diagonal. The expressions of their submatrices can be found in [20, page 3]. The function f_k^2 is unknown, as well as the variance of v_k^2 . As for the observation model, it is given by (1), but by adding a white noise w_k that encompasses the uncertainties in the measurement process:

$$y_k = h_k(\mathbf{x}_k) + w_k, \quad (4)$$

where the actual PEMFC power writes $P_k = h_k(\mathbf{x}_k)$. A difficulty is that the measurement noise w_k is non-Gaussian, since it is degraded by outliers. Its variance is denoted Ω_k .

C. Proposed hybrid LSTM and robust iterative KF

A straightforward application of an EKF using the state space representation (2)-(4) cannot be considered as f_k^2 is unknown and the measurement noise is non-Gaussian. The contribution of the paper is twofold: firstly, an LSTM network is integrated at the filter prediction step to learn the dynamic behaviour of Γ_k , and secondly, a robust loss function is considered for the LSTM training in combination with a robust version of the EKF. It results in a hybrid algorithm that tightly combines expert knowledge and deep-learning techniques.

LSTMs are a class of recurrent NNs which are efficient to capture long-term dependencies and transient responses that may not be explicitly modelled [21]. Widely employed for automatic speech translation and time-series processing in general, their architecture makes it possible to generate outputs thanks to sequential combinations of inputs and previous outputs [22]. They can handle variable time-series length and

have the advantage of mitigating gradient vanishing during training.

The proposed approach is illustrated on Fig 1. At each time step k , the prediction step leverages two LSTM NNs: \mathcal{L}_x that is dedicated to the prediction of x_k^2 , and \mathcal{L}_P that propagates its posterior variance but also its intercorrelations with the well-modelled states \mathbf{x}_k^1 . In the sequel, the index notation $k|l$ combined with an upper hat refers to the state estimation at time k by taking into account all the measurements up to time l . The hybrid prediction writes:

$$\hat{\mathbf{x}}_{k|k-1}^1 = F_k^1 \hat{\mathbf{x}}_{k-1|k-1}^1, \quad (5)$$

$$(\hat{x}_{k|k-1}^2, \mathbf{m}_k^x) = \mathcal{L}_x(\hat{x}_{k-1|k-1}^2, \mathbf{m}_{k-1}^x), \quad (6)$$

where \mathbf{m}_k^x gathers the hidden and cell states of the LSTM NN at time k . The posterior covariance matrix of the state, denoted $P_{k|l}$ when based on the measurements up to time l , is also updated by block. By denoting $P_{k|l}^1$ the posterior covariance matrix of \mathbf{x}_k^1 , it ensues:

$$P_{k|k-1}^1 = F_k^1 P_{k-1|k-1}^1 (F_k^1)^\top + Q_k^1. \quad (7)$$

In parallel, a second LSTM NN is leveraged for the remaining blocks of $P_{k|k}$:

$$\begin{aligned} & (P_{k|k-1}^2, P_{k|k-1}^{1,2}, \mathbf{m}_k^P) \\ & = \mathcal{L}_P(P_{k-1|k-1}^2, P_{k-1|k-1}^{1,2}, \mathbf{m}_{k-1}^P, \hat{\mathbf{x}}_{k-1|k-1}^1), \end{aligned} \quad (8)$$

with \mathbf{m}_k^P containing the hidden and cell states of the latter. $P_{k|l}^2$ and $P_{k|l}^{1,2}$ stand for the posterior variance of x_k^2 and the posterior intercovariance matrix of \mathbf{x}_k^1 and x_k^2 , respectively.

After state propagation, the estimates are updated by taking into account the current power measurement. Robust KFs consist in re-interpreting this step as the minimization of a cost function and replacing the squared L_2 -norm by a robust one. A Huber norm that mitigates the effects of abrupt changes in the observation noise [23], [24] is usually considered. $\hat{\mathbf{x}}_{k|k}$ is obtained by minimizing the following criterion [18]:

$$\mathcal{J}(\mathbf{x}) = \|\mathbf{x} - \hat{\mathbf{x}}_{k|k-1}\|_{P_{k|k-1}}^2 + R_\delta \left(\Omega_k^{-\frac{1}{2}} (y_k - h_k(\mathbf{x})) \right) \quad (9)$$

with $\Omega_k^{-1/2}$ the square-root of Ω_k^{-1} . The function R_δ denotes the Huber norm [24].

A Gauss-Newton (GN) algorithm is usually applied to minimize (9) and the posterior covariance $P_{k|k}$ comes as a Gauss-Laplace approximaton. The resulting algorithm is referred to

as robust iterated KF (IKF). It should be noted that only a few iterations are required in practice since the GN optimizer is embedded in a KF. In the considered setting, each iteration is equivalent to applying a classical EKF correction step but refining iteratively the linearization point and weighting the measurement noise covariance matrix. The l^{th} GN iteration can be written:

$$K_k^{(l+1)} = P_{k|k}^{(l)} (H_k^{(l)})^\top \left(H_k^{(l)} P_{k|k}^{(l)} (H_k^{(l)})^\top + \tilde{\Omega}_k^{(l)} \right)^{-1} \quad (10)$$

$$\mathbf{x}_{k|k}^{(l+1)} = \mathbf{x}_{k|k-1} + K_k^{(l+1)} (y_k - \mathbf{h}_k(\mathbf{x}_{k|k}^{(l)})). \quad (11)$$

In the latter expression, $H_k^{(l)}$ is the Jacobian of \mathbf{h}_k computed at $\mathbf{x}_{k|k}^{(l)}$ and $\tilde{\Omega}_k^{(l)}$ is obtained as:

$$\tilde{\Omega}_k^{(l)} = \Omega_k^{1/2} (W_k^{(l)})^{-1} \Omega_k^{1/2} \quad (12)$$

with $W_k^{(l)}$ a weighting matrix expressed as

$$W_k^{(l)} = \text{diag} \left[w \left(\Omega_k^{-1/2} \left(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_{k|k}^{(l)}) \right) \right) \right]. \quad (13)$$

The weight function $w(\cdot)$ is derived from the Huber score function. Finally, the covariance is updated by:

$$P_{k|k}^{(l+1)} = (I - K_k^{(l+1)} H_k^{(l)}) P_{k|k}^{(l)}. \quad (14)$$

For the initialization, $\mathbf{x}_{k|k}^{(0)} = \hat{\mathbf{x}}_{k|k-1}$ and $P_{k|k}^{(0)} = P_{k|k-1}$. The final posterior state estimate $\hat{\mathbf{x}}_{k|k}$ and covariance matrix $P_{k|k}$ are taken as the outputs of the last GN iteration.

D. Training step

The training aims to adjust the parameters θ of the two LSTM NNs so as to yield an estimated PEMFC power as close as possible to the actual one during a reference time period. It is classically performed by minimizing a cost function \mathcal{C} . However, the outages in the measurements can also impair this crucial step. To prevent such difficulties, we propose to consider also a Huber function [24]:

$$\mathcal{C}(\theta) = \frac{1}{N} \sum_{k=1}^N R_\delta(y_k - \hat{y}_{k|k-1}), \quad (15)$$

where $\hat{y}_{k|k-1} = \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1})$.

The whole procedure is the following: the proposed hybrid IKF is run for N time steps which constitute the training period and the successive estimates are stored and used to infer the PEMFC power predictions $\hat{y}_{k|k-1}$ up to $k = N$. The gradients of the network parameters are then computed by back-propagation and are used to update the weights and biases of the networks. These computations are reiterated over the training period until a stopping criterion is satisfied.

III. EXPERIMENTS

The dataset corresponds to a 1100 h test [25] of a 5-cell PEMFC stack with an active area of 100 cm^2 . To keep only the part that aged at constant current, the data are shortened at 985 h. The nominal current density is 0.70 A cm^{-2} and the nominal temperature is 55°C . Intensities and voltages are combined to generate single power measurements. The power

critical threshold is equal to 96% of the initial power [26] and it is reached at 884.5 h. In the sequel the prognosis starts at 325 h which corresponds to a RUL of 559.5 h.

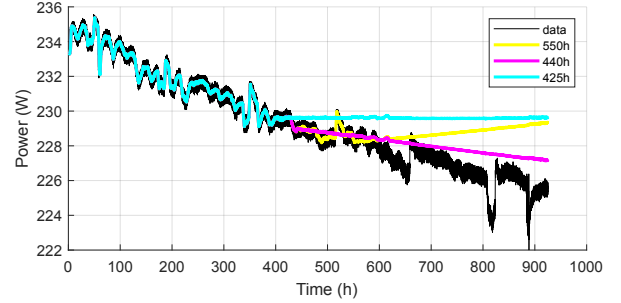


Fig. 2. Examples of AEKF prognostic after different training times.

The proposed algorithm (denoted Algo 1 hereafter) is confronted with alternative approaches either from the literature or variants of its architecture to show the interest of both the tight hybridization and the use of robust cost functions. The non robust version of our algorithm (Algo 2), a simple hybrid LSTM-EKF (Algo 3) wherein the entire prediction step is carried out by a LSTM NN as presented in [6], and the robust version (Algo 4) of the latter are considered. It should be noted that the latter has not been published and is a by-product of our study. A comparison is also made with a knowledge-based method using an AEKF [14].

As for implementation, considering Algo 1 and Algo 2, one LSTM layer with fifty neurons for x_k^2 and three LSTM layers with forty neurons each for $P_{k|k-1}^2$ and $P_{k|k-1}^{1,2}$ are trained. For Algo 3 and Algo 4, covariances are fixed and three LSTM layers with thirty neurons each are used to estimate all of \mathbf{x}_k . Weights and biases are updated with an AdamW optimizer [27]. For the robust counterparts, at each time step, five iterations of the GN algorithm are performed for the robust KF updating. The model is trained during the 325 first hours. The considered stopping criterion is cost function stabilization.

Firstly, the AEKF is tested to estimate RUL. It appears that the long-term degradation trend captured by this algorithm highly depends on the local bending of the signal at the start of the prognostic, as illustrated in Fig. 2. This sensitivity discards it to accurately estimate an RUL.

For hybrid algorithms, the parallel is made thanks to the accuracy of the RUL prediction and the Mean Absolute Percentage Error criterion of the PEMCF power estimation: $\text{MAPE} = \frac{1}{N-n} \sum_{k=n}^N |(y_k - \hat{y}_{k|k-1})/y_k|$ with n corresponding to 325 h. This indicator allows us to verify the accuracy of the long-term forecasting. The average of ten training sessions with varying initializations led to the results depicted in table I. It should be noted that a RUL underestimation leads to negative error. It can be observed that our method outperforms the others in terms of accuracy of both the RUL and power degradation estimations. A major issue with Algo 3 and Algo 4 is that the neural networks experience difficulty in discriminating signal and noise. Furthermore, whatever the

hybrid architecture, the robust versions yield the best results. This further reinforces the importance of our model when confronted with data showing significant variations. Fig. 3 illustrates MAPE differences between the two best algorithms compared to database.

TABLE I
ALGORITHM COMPARISON

Hybrid IKF-LSTM Model	Criteria		
	RUL	RUL error	MAPE
Algo 1	562.2 h	2.7 h	0.001 81
Algo 2	306.0 h	-253.5 h	0.010 24
Algo 3	584.9 h	25.4 h	0.004 02
Algo 4	549.1 h	-10.4 h	0.002 73

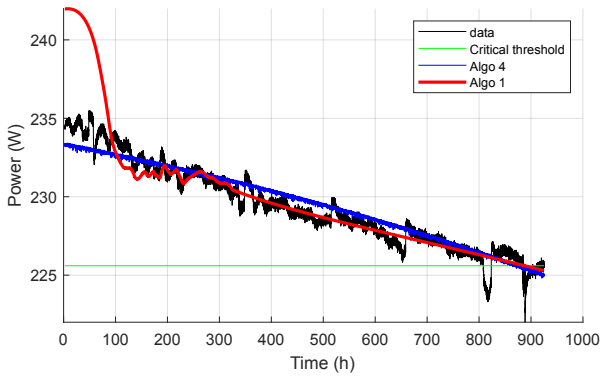


Fig. 3. Comparison of the two best forecasting algorithms.

Finally, another contribution of our work is the reduction in the required training time to estimate an RUL. Compared to the literature, like [14] or [15] that have a minimum training set of 55 % to 60 % with an RUL estimation error of 3 hours, we propose a slightly more precise RUL estimation (error of 2.7 hours) with only 32.5 % of the training set.

IV. CONCLUSION AND PERSPECTIVES

This paper proposes a novel approach to perform prognostics of fuel cell stacks by RUL estimation. It tightly fuses a robust KF with deep learning techniques. The latter interact with classical evolution models so that they only learn the dynamics of the parameters that are difficult to characterize. A good accuracy is obtained when testing the algorithm to predict PEMFC power from a reference database. In perspective, it would be interesting to study alternative hybridizations involving transformers and/or particle filters to better handle the non linear measurement model.

REFERENCES

- [1] R. K. Mobley, *An Introduction to Predictive Maintenance (Second Edition)*, Plant Engineering. Butterworth-Heinemann, Burlington, second edition edition, 2002.
- [2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [3] P. Del Moral, "Non linear filtering: Interacting particle solution," *Markov Processes and Related Fields*, vol. 2, pp. 555–580, 03 1996.
- [4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

- [5] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.
- [6] Z. Zhang, Y.-X. Wang, H. He, and F. Sun, "A short and long-term prognostic associating with remaining useful life estimation for proton exchange membrane fuel cell," *Applied Energy*, vol. 304, pp. 117841, 2021.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.
- [8] W. Qin, J. Tang, C. Lu, and S. Lao, "Trajectory prediction based on long short-term memory network and kalman filter using hurricanes as an example," *Computational Geosciences*, vol. 25, pp. 1–19, 06 2021.
- [9] R. Xu, Y. Wang, and Z. Chen, "A hybrid approach to predict battery health combined with attention-based transformer and online correction," *Journal of Energy Storage*, vol. 65, pp. 107365, 2023.
- [10] E.A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, 2000, pp. 153–158.
- [11] H. Coskun, F. Achilles, R. DiPietro, N. Navab, and F. Tombari, "Long short-term memory Kalman filters: Recurrent neural estimators for pose regularization," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [12] J. Kim and K. Lee, "Unscented kalman filter-aided long short-term memory approach for wind nowcasting," *Aerospace*, vol. 8, 08 2021.
- [13] Ziqiao Zhang, Mengxue Hou, Fumin Zhang, and Catherine R. Edwards, "An lstm based kalman filter for spatio-temporal ocean currents assimilation," in *Proceedings of the 14th International Conference on Underwater Networks & Systems*, New York, NY, USA, 2020, WUWNet '19, Association for Computing Machinery.
- [14] R. Pan, D. Yang, Y. Wang, and Z. Chen, "Performance degradation prediction of proton exchange membrane fuel cell using a hybrid prognostic approach," *International Journal of Hydrogen Energy*, vol. 45, pp. 30994–31008, 11 2020.
- [15] Z. Xia, Y. Wang, L. Ma, Y. Zhu, Y. Li, J. Tao, and G. Tian, "A hybrid prognostic method for proton-exchange-membrane fuel cell with decomposition forecasting framework based on aekf and lstm," *Sensors*, vol. 23, no. 1, 2023.
- [16] X. Zhang, Y. Huang, Z. Zhang, H. Lin, Y. Zeng, and M. Gao, "A hybrid method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network combined with attention and a kalman filter," *Energies*, vol. 15, pp. 6745, 09 2022.
- [17] R. Ma, R. Xie, L. Xu, Y. Huangfu, and Y. Li, "A hybrid prognostic method for pemfc with aging parameter prediction," *IEEE Transactions on Transportation Electrification*, vol. PP, pp. 1–1, 04 2021.
- [18] D. Medina, H. Li, J. Vilà-Valls, and P. Closas, "Robust filtering techniques for RTK positioning in harsh propagation environments," *Sensors*, vol. 21, no. 4, 2021.
- [19] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, and N. Zerhouni, "PEMFC ageing modeling for prognostics and health assessment," *9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS 2015)*, vol. 48, no. 21, pp. 790–795, 2015.
- [20] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. part i: Dynamic models," *IEEE Trans. on Aerospace and Elec. Syst.*, vol. 39, pp. 1333–1364, 2003.
- [21] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 5695–5705, 2018.
- [22] J. L. McClelland and D. E. Rumelhart, *Schemata and Sequential Thought Processes in PDP Models*, pp. 7–57, MIT press, 1987.
- [23] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, 2005.
- [24] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73 – 101, 1964.
- [25] F. Harel, "Défi IEEE PHM data 2014. FCLAB," 2014.
- [26] M. Jouin, R. Gouriveau, D. Hissel, M. C. Péra, and N. Zerhouni, "Prognostics of PEM fuel cell in a particle filtering framework," *International Journal of Hydrogen Energy*, vol. 39, no. 1, pp. 481–494, 2014.
- [27] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.