A Probabilistic Data-Driven Prognostic Methodology for Proton Exchange Membrane Fuel Cells

Amrit Sethi[1], and Dries Verstraete[2]

[1,2] *School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney*
amrit.sethi@sydney.edu.au
dries.verstraete@sydney.edu.au

ABSTRACT

Hydrogen fuel cells, particularly proton exchange membrane fuel cells (PEMFC), are promising, robust, clean energy sources. However, their high cost and short lifespan under dynamic loads impedes their widespread usage. Accurate and real-time prognostics, especially remaining useful time (RUL) estimation, can help ameliorate the commercial viability of PEMFCs. Data-driven methods are increasingly considered for RUL estimation. This paper looks at two such methods – Gaussian Process Regression (GPR) and Long-Short Term Memory (LSTM) Networks and assesses them in terms of accuracy and suitability for real-time applications when tested against the IEEE PHM Challenge 2014 data set. Gaussian Process Regression is a non-parametric kernel method. LSTM, on the other hand, is a recurrent neural network based architecture that is effective at detecting both long term and short term trends in time series predictions. For the cases investigated here, the results derived using LSTM are more accurate, especially since they effectively capture long term trends. However, GPR assigns a probability to its prediction - a desirable aspect in a real-time setting so that corrective action can be applied appropriately. The paper then proposes the use of a variant of these methods - Gaussian Process-Long Short Term Memory Network (GP-LSTM) as an alternative that combines the higher accuracies of the LSTM method and the probabilistic output from GPR. The results attained using GP-LSTM are close in accuracy to the LSTM results and have a probability associated with them, making them suitable for real-time applications. The effectiveness of GP-LSTM is further proven using a dynamic data set and strategies are suggested to appropriately apply GP-LSTM to real-world scenarios.

1. INTRODUCTION

Fuel cells have seen a lot of growth over the recent years. This development results as the need to address the consequences of electricity production and vehicle propulsion using fossil fuels becomes more prominent. Proton exchange membrane fuel cells (PEMFCs), in particular, are found to be of considerable interest due to their low operating temperature, robustness, and high energy density (Larminie & Dicks, 2001). Owing to such advantages, they have found uses in many automotive and aerospace applications (Sharaf & Orhan, 2014; Gong & Verstraete, 2017; Gong, Palmer, Brian, Harvey, & Verstraete, 2016; Verstraete, Gong, Lu, & Palmer, 2015; Verstraete et al., 2014). However, their short lifespans do not justify their high cost, warranting the need for effective prognostics (Sutharssan et al., 2017; Sharaf & Orhan, 2014; H. Liu, Chen, et al., 2020). Many studies relating to the prognostics of PEMFCs, and in particular remaining useful life (RUL) estimation, have been carried out. The bulk of this research can be divided into three main categories - (i) model-driven methods, (ii) data-driven methods and (iii) fusion methods which combine elements of model-driven and data-driven methods (Sutharssan et al., 2017; H. Liu, Chen, et al., 2020).

Model-driven approaches involve tracking the underlying degradation mechanisms by using empirically derived models, the physics associated with the failure mode, or a combination of both (Sutharssan et al., 2017; H. Liu, Chen, et al., 2020). Filter based methods are often combined with PEMFC models for state estimation and prediction. For example, (Bressel, Hilairet, Hissel, & Ould Bouamama, 2016) use a polarisation voltage loss model where certain parameter trends over time are established empirically and an extended Kalman filter is applied for state estimation and prediction. (Jouin, Gouriveau, Hissel, Péra, & Zerhouni, 2014) apply a particle filter in conjunction with an empirical state transition model to determine voltage degradation as well as to estimate RUL. Similarly, Kimotho et al. (Kimotho, Meyer, & Sextro, 2015) also examine voltage degradation using particle filters, but additionally include the use of a self-healing

factor to account for boosts in voltage after characterisation tests. Jha et al. (Jha, Bressel, Ould-Bouamama, & Dauphin-Tanguy, 2016) derive a bond graph model of a PEMFC and use a particle filter for diagnostics and prognostics. While model-driven methods are prevalent, they can be quite complex (in terms of design and computation) and may require a strong understanding of underlying degradation mechanisms (H. Liu, Chen, et al., 2020). This has led to data-driven methods becoming increasingly popular for prognostics.

Data-driven methods are often preferred due to their ease of implementation as well as their ability to compute the required output rapidly - a desirable feature for online diagnosis and prognostics. (L. Zhu & Chen, 2018) propose the use of Gaussian Process State Space Model where Gaussian Processes are used for modelling state transitions for a state space framework and to track and predict voltage degradation. The model provides a probabilistic output that has the added advantage of providing a confidence interval for the associated predictions. (Wu, Breaz, Gao, Paire, & Miraoui, 2016) use an adaptive relevance vector machine for prognostics. Research in neural-network-based architectures has also resulted in their advent in PEMFC prognostics (H. Liu, Chen, et al., 2020; H. Liu, Chen, Hissel, & Su, 2019). Javed et al. (Javed, Gouriveau, Zerhouni, & Hissel, 2016b, 2016a) use summation wavelet - extreme learning machines in addition to an incremental learning approach for accurate predictions involving variable loads. (Chen, Laghrouche, & Djerdir, 2019) use wavelet analysis for decomposing the signal as an input for their extreme learning machine- genetic algorithm based degradation model. (H. Liu, Chen, Hou, Shao, & Su, 2017) also used wavelet analysis for breaking the signal down to its different frequency components and apply group method of data handling for short term prognostics. Recurrent neural networks and their derivative architectures - primarily long short term memory networks (LSTM networks) have also been employed because of their ability to handle sequential data.

(J. Liu et al., 2019) apply an LSTM network for remaining useful life (RUL) estimation and obtain results with much higher accuracies than traditional backpropagation neural networks. Yang et al. integrate LSTMs and use them to derive predictions that are then passed into particle filters for higher precision (Yang et al., 2017). (Ma, Breaz, et al., 2018) use LSTM networks for different fuel cells and operating conditions and obtain accurate results for both - long term and short term predictions. They then use the Grid LSTM network (Ma, Yang, et al., 2018) - a modification of the traditional LSTM cell that allows for higher accuracies. Furthermore, they combine LSTM networks with the auto-regressive integrated moving average (ARIMA) method to provide accurate predictions in addition to accounting for fuel cell performance recovery (Ma et al., 2019). (Wang, Cheng, & Hsiao, 2020) apply stacked long short term networks with hyperparameters derived using the differential evolutionary algorithm to achieve even higher accuracies in predictions. Despite a breadth of research in PEMFC prognostics methodologies, there is insufficient exploration of the feasibility of applying such methods in an *online* prognostic framework where it can be deployed for real-time (or close to real-time) operation. Furthermore, most data-driven methods do not provide uncertainty measurements like many of their model-based or hybrid counterparts. Uncertainty related information can be beneficial for an online PHM framework as confidence in the measurements can be used to help determine the extent of corrective action required (Boškoski, Debenjak, & Boshkoska, 2017; Jouin, Gouriveau, Hissel, Péra, & Zerhouni, 2013).

This study first explores Gaussian Process Regression (GPR) and LSTM networks in terms of prediction accuracy and suitability for online diagnostics. The suitability for online operation is determined by establishing 'pseudo-online' conditions where we control the amount of past data the model has access to, similar to a model operation in a real-world application. The study then proposes the use of GP-LSTM (Al-Shedivat, Wilson, Saatchi, Hu, & Xing, 2017) which combines the advantages of both methods by learning the trends of the degradation through LSTMs and being able to provide an uncertainty associated with the prediction to make it more suitable for online prognostics of PEMFCs. GP-LSTM provides an uncertainty measurement for the predictions directly due to the nature of Gaussian Processes. It is also more suited for an online context when compared to other methods, such as the one in (Yang et al., 2017) which uses a particle filter - LSTM fusion since GP-LSTM does not require multiple predictions to determine each output value. Primarily, GP-LSTMs can provide uncertainty for predictions at a significantly lower computational cost when compared to other LSTM based methodologies (Al-Shedivat et al., 2017). Even though GP-LSTM has been been used for time-series problems and sequential data because of the advantages it offers (Qin, Zhu, Qin, & Wang, 2019; S. Zhu, Yuan, Xu, Luo, & Zhang, 2019), it has not yet been explored for fuel cell prognostics. This study further explores the suitability of GP-LSTM and the aforementioned methods for real-world applications by comparing their performance in an autoregressive setting as well as a 'free simulation' setting - where the model uses its outputs as inputs for future predictions. GP-LSTMs are also explored for prognostics involving dynamic conditions.

2. PEMFC SYSTEM

This study uses the IEEE PHM Data Challenge (FCLAB, 2014) provided by the FCLAB research federation for analysis when employing the different methodologies. The data set consists of ageing and health data from two 5-cell stacks - FC1 and FC2. The cells have an active area of 100 cm^2 and were subjected to a nominal current of 0.70 A/cm^2. FC1 was

operated at constant current load and FC2 was subjected to dynamic conditions - a steady current with ripples. FC1 and its associated data will be the main focus of this study. However, GP-LSTM, with modifications and retraining, will be applied to data from FC2 to prove that the method can be applied to more dynamic profiles. During the test, numerous parameters were recorded including the stack temperature, gas flow and reactant hygrometry rates, cell and stack voltages, as well as the operating current. In addition, characterization tests in the form of electrochemical impedance spectroscopy (EIS) were also run periodically.

To suitably perform prognostics and remaining useful life (RUL) estimation, an appropriate ageing characteristic needs to be chosen. The stack power is a defining feature for stack degradation and is used as the ageing parameter (Jouin, Gouriveau, Hissel, Péra, & Zerhouni, 2015). The power can be computed by taking the product of the current and stack voltage. The power output during the ageing test is shown in Figure 1. The test was run for 1,154 hours and 143,862 points were generated. To make the predictive model suitable for online diagnostic and updating, the input variables were restricted to past time and power values. The existing data set was trimmed by applying a median filter with a one hour window over the data to obtain a new data set of 1,154 points that helps remove large spikes and noise (Kimotho et al., 2015). A sliding window of 10 points is used for performing predictions (Al-Shedivat et al., 2017). The first 670 hours are used for training and the rest is the test data, as can be seen in Figure 1. To effectively test the models for suitability in an online context, where some data may not be available when the model is operating, two different forms of predictions were made (Al-Shedivat et al., 2017):

- **Autoregression (AR)** - One-step-ahead prediction is performed. True past power values are used for predicting the next power value. (Zhou, Gao, Breaz, Ravey, & Miraoui, 2017; Cheng, Zerhouni, & Lu, 2018; Al-Shedivat et al., 2017)
- **Free Simulation (FS)** - The model only uses the true power values from the training set. In the test set, the model maps its predictions as inputs for future predictions. This is suitable for determining the general trend of the stack power and for performing RUL determination well ahead of time (Al-Shedivat et al., 2017). This helps emulate online conditions where the model may have to predict power values far in the future instead of just the power at the next time step.

To compare the efficacy of all the models for both prediction forms, the power threshold for RUL estimation is set at 3.5% of the initial power value, or 226.87W (FCLAB, 2014). The root-mean-square error (RMSE) and mean absolute percentage error (MAPE) are then used to further compare the tracking capabilities of all the models in an autoregressive context and can be found using (H. Liu et al., 2019):

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

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{|y_i|}$$

where y_i and $\hat{y}_i$ represent the actual and predicted values respectively, N represents the number of points used in the measure - the number of test points used in this study.

3. RUL ESTIMATION METHODS

This study explores the use of long short term memory networks and Gaussian process regression for fuel cell prognostics and the methods are briefly described. The study then proposes the use of GP-LSTM for PEMFC prognostics, which combines the two aforementioned methods and explains how this method retains the advantages of LSTM networks and Gaussian process regression.

3.1. Recurrent Neural Network and Long Short Term Network

A recurrent neural network (RNN) is a form of neural network designed to deal with sequential data and has found a myriad of applications for time series problems. In a RNN, an inner loop is used to loop back previous information within the hidden layers to account for the past state, in conjunction with the current input to form a prediction (Al-Shedivat et al., 2017). While useful, RNNs cannot deal with long term trends effectively due to the vanishing gradient problem where the impact of backpropagation during neuron weight updates diminishes exponentially over the different time steps. To account for this, long short term memory networks (LSTM) are popularly used (Wang et al., 2020; Géron, 2019).

LSTMs use a more elaborate structure as shown in Figure 2 (Géron, 2019). Here x_t represents the current input, h_t the current output and c_t represents the current cell state. $\oplus$ is the addition operation and $\otimes$ is element-wise multiplication. σ and $\tanh$ are found using:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$\tanh(z) = 2\sigma(2z) - 1$$

The LSTM cell structure is comprised of different gates and layers which control the amount of historical information that is preserved and how the historical information is used in conjunction with the input to determine the output (Géron, 2019):

- **Forget Gate** - This gate controls the information that can be 'forgotten' from the previous cell state c_{t-1} - fun-

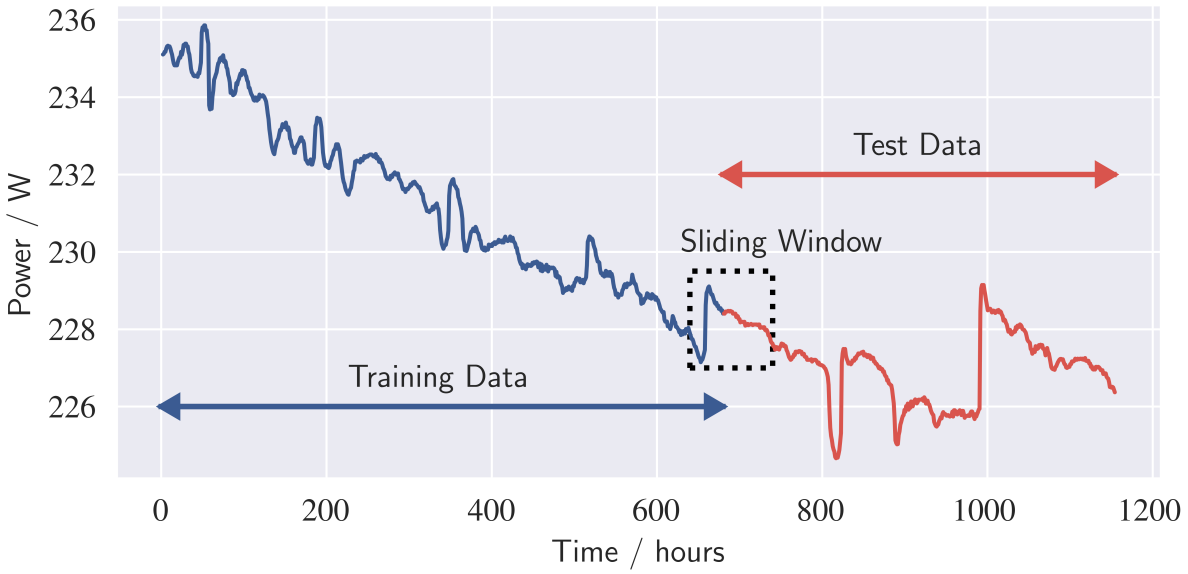


Figure 1. Stack performance over time measured using the power output

damentally controlling the components of the long term state that are erased. The output of f_t can be computed using:

$$f_t = \sigma(W_{xf}^\top x_t + W_{hf}^\top h_{t-1} + b_f)$$

where the generic form $W_{\alpha\beta}$ represents the weight matrices between connections α and β and b_m represents the bias for output m. For example, W_{xf} represents the weight matrix between the input x and the forget gate f, and b_f represents the bias for the output of the forget gate.

- **Input Gate** - g_t analyses the current input x_t and the previous cell output h_{t-1}. The input gate i_t controls which components of g_t are preserved:

$$i_t = \sigma(W_{xi}^\top x_t + W_{hi}^\top h_{t-1} + b_i)$$
$$g_t = \tanh(W_{xg}^\top x_t + W_{hg}^\top h_{t-1} + b_g)$$

- **Output Gate** - This controls the parts of the cell state that are preserved to determine the output h_t:

$$o_t = \sigma(W_{xo}^\top x_t + W_{ho}^\top h_{t-1} + b_o)$$
$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t$$
$$h_t = o_t \otimes \tanh c_t$$

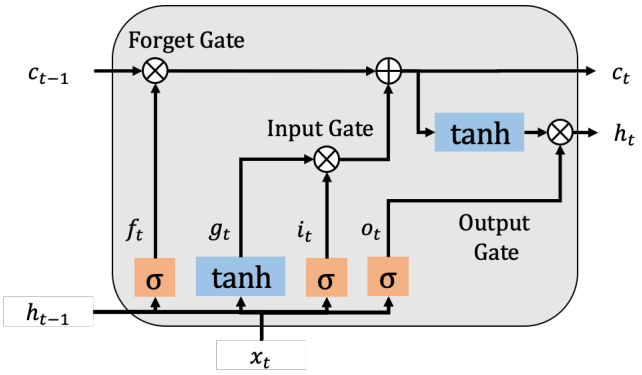


Figure 2. LSTM Cell Structure

3.2. Gaussian Process Regression

Gaussian process Regression (GPR) is a supervised learning method that uses a Gaussian distribution over functions to describe the feature space (Murphy, 2014). This Gaussian distribution can be described using its mean function $m(\mathbf{x})$ and covariance function $\kappa(\mathbf{x}, \mathbf{x}')$.

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), \kappa(\mathbf{x}, \mathbf{x}'))$$

Essentially, for a finite number of inputs, $[\mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_n}]$, the distribution over the functions can better be understood as a joint Gaussian distribution over the function evaluations

$$p(f(\mathbf{x_1}), f(\mathbf{x_2}), ...f(\mathbf{x_n})) = p(\mathbf{f}|\mathbf{x})$$
$$= \mathcal{N}(\boldsymbol{\mu}|\mathbf{K})$$

where $\mathbf{K}$ is a matrix with the terms $K_{i,j} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ and $\boldsymbol{\mu} = (m(\mathbf{x}_1, ...m(\mathbf{x}_N))$

Using this definition, training inputs $\mathbf{x}$, training targets $\mathbf{y}$, the posterior distribution over function evaluations $\mathbf{f}_*$ for a test input $\mathbf{x}_*$ can be described as (Murphy, 2014; Richardson, Osborne, & Howey, 2019):

$$p(\mathbf{f}_*|\mathbf{x}_*, \mathbf{x}, \mathbf{y}) = \mathcal{N}(\mathbf{f}_*|\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$$

where

$$\boldsymbol{\mu}_* = \mathbf{K_{x_*x}}^\top \mathbf{K_{xx}}^{-1}\mathbf{y}$$
$$\boldsymbol{\Sigma}_* = \mathbf{K_{x_*x_*}} - \mathbf{K_{xx_*}}^\top \mathbf{K_{xx}}^{-1}\mathbf{K_{xx_*}}$$

Gaussian Process Regression is commonly used in time series modelling and the whole training set is typically used for performing predictions. To allow the model to scale effectively, modifications can be introduced, such as active learning (Roberts et al., 2013), whereby the training set is replaced by a subset of the existing points that best describe the feature space or variational learning, where the existing GPR model is replaced by a computationally less expensive model by finding key inducing points in the feature space (H. Liu, Ong, Shen, & Cai, 2020). However, in this study we employ the whole data set and also apply a sliding window to gauge the effectiveness of GPRs in determining short term and long term trends. To predict the power value using a sliding window of size L, an L number of previous predictions are used as features in the regression model.

For implementing GPR time series modelling in this study, a mean function of $m(\mathbf{x}) = 0$ suffices as GPs are quite flexible (Murphy, 2014). A variety of kernel functions are used in practice with the most popular one for times series prediction being the squared exponential kernel - the one used in this study (Roberts et al., 2013). The squared exponential kernel describes the covariance between two inputs, x_1 and x_2, as follows:

$$\kappa(x_1, x_2) = \sigma_f^2 exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Thus, the influence two input points have over their covariance wanes with distance l. In fact, most kernel functions used in practice share this attribute. For time series prediction, this ends up affecting the quality of the predictions as the model may fail to capture long term trends effectively. This is where GP-LSTMs prove to be quite effective.

3.3. GP-LSTM

GP-LSTMs combine the long term effectiveness of LSTMs with the probabilistic non-parametric regression methodology of GPs (Al-Shedivat et al., 2017). This is achieved by employing LSTM based kernels in a Gaussian Process framework. By doing so, the kernel can hold long term trend knowledge gained by the nature of the LSTM and use the GP method to determine the uncertainty associated with this method when employing the LSTM based kernel. The reader is directed to (Al-Shedivat et al., 2017) for a more comprehensive explanation.

4. RESULTS AND DISCUSSION

Gaussian Process Regression and Long Short Term Memory Networks are first applied and compared in terms of their predictive performance for autoregression and free simulation. Then, GP-LSTM is applied and demonstrated as a suitable alternative to the aforementioned techniques due to its ability to learn long term trends as well as provide probabilistic estimates. GP-LSTM is further explored for determining its suitability in an online context by using a combination of autoregressive and free simulation predictions.

4.1. Gaussian Process Regression

Initially, the whole training data set is used for both autoregression and free simulation. The free simulation result is shown in Figure 3. Although the model fits the training data well and disparities are almost imperceptible, the fit over the test data is a completely different story. The model predicts the initial points well but stops doing so after 10-15 hours. It appears to emphasise the short term trends over the long term, and ends up propagating this short term trend through the entirety of the prediction time. As a result, the predictions never reach the RUL threshold. The uncertainty associated with the predictions increases rapidly - as expected given the nature of the squared exponential kernel. Even though most of the actual data points are within the confidence interval of the predictions, the uncertainty is too high to be of any practical use.

Since it might not be feasible to use all of the training data in an online context, the sensitivity of the model was tested by applying a sliding window for training and testing. The predictions associated with applying a sliding window are shown in Figure 4. The highest uncertainties in the predictions are generally associated with the points following large transitions, such as the ones around 900 hours and 1000 hours. The free simulation predictions are shown in Figure 4b. The model is able to predict the initial trend but deviates pretty rapidly and flattens out after approximately 50 hours of prediction. The End-of-Life (EOL) estimates associated with the different methods can also be seen in Table 2. As expected based on Figure 4a, the autoregression results are very accurate but the free simulation results are not. The GPR model in free simulation heavily underestimates the end of life (EOL) - a testimony to its inability to suitably track the long term trend associated with the stack degradation. Therefore, LSTMs are explored to remedy this drawback.

4.2. Long Short Term Memory (LSTM) Network

LSTM Networks are designed to capture both long term and short term trends for time-series predictions. This functionality is quite apparent in Figure 5. The autoregression and free simulation results were obtained using the same model - a single LSTM layer with 50 units, $\tanh$ as the activation function

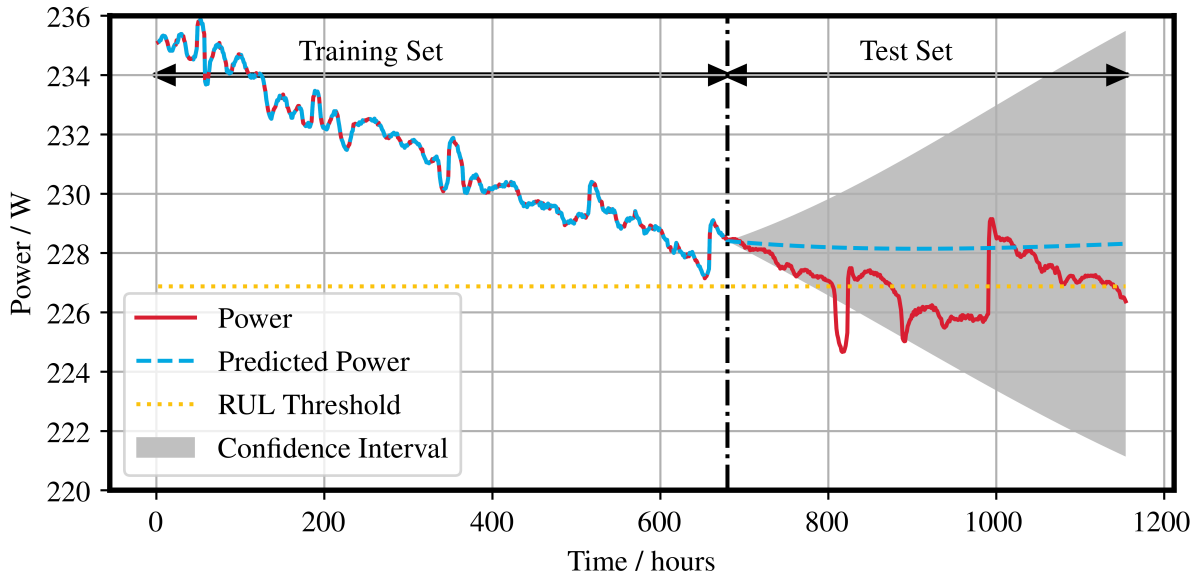
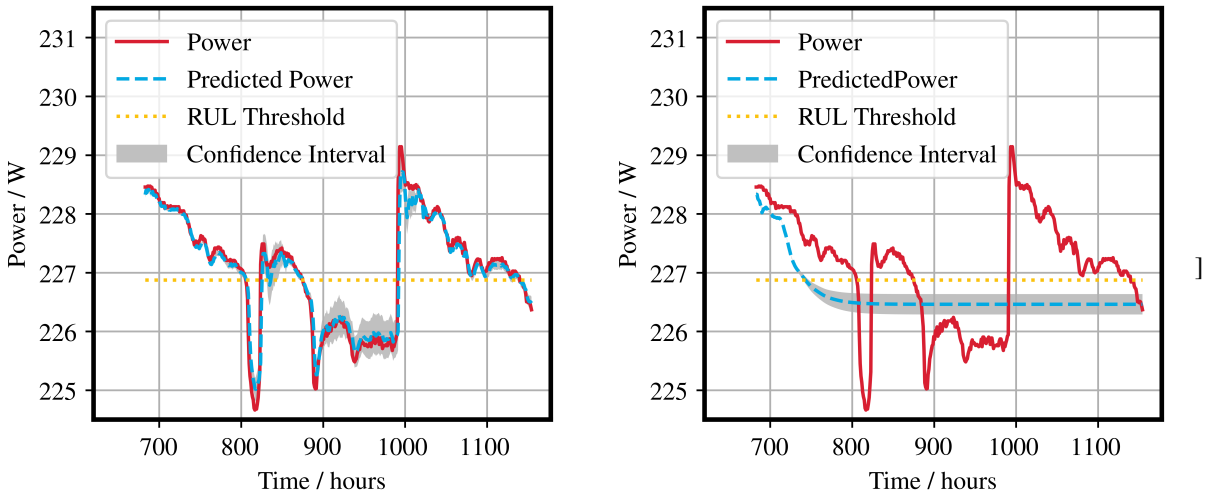


Figure 3. Applying GPR over the whole training data set



(a) GPR Autoregression

(b) GPR Free Simulation

Figure 4. GPR using a sliding window

Table 1. RMSE and MAPE values associated with the autoregression predictions from the different models

Model	RMSE	MAPE
GPR	0.2261	0.2258
LSTM	0.3125	0.3092
GP-LSTM	0.3510	0.3156

and a dropout and recurrent dropout of 0.05 and 0.1 respectively. The hyperparameters are selected based on values provided by other LSTM related papers (J. Liu et al., 2019; Wang et al., 2020) and on heuristics suggested by (Géron, 2019) to reduce over-fitting by changing the respective dropout rates. Figure 5a shows the autoregression results which fit the actual power values quite well. The model does under-predict the extent of the transitions when they are relatively drastic compared to surrounding points, such as the ones around 810 hours, 890 hours and 1000 hours. However, in most cases, it ends up converging with the actual values soon after in a corrective manner. The RMSE and MAPE values associated with the autoregression predictions are listed in Table 1 and are higher than those of GPR. The higher accuracy linked to the GPR predictions can be attributed to the inherent nature of GPR modelling which involves using pairwise trends between adjacent points through kernels (Al-Shedivat et al., 2017)

However, LSTM's long term trend recognition in the free simulation setting is far more accurate, as manifested in Figure 5b. The mode appears to follow the downward trend and has an understanding of the "bounce back" in the trend - the sudden increase around 1000 hours caused by the frequent characterisation tests. The final power prediction at 1,154 hours is quite close to the real power value. However, the initial modelling of the downward trend over-predicts for a while before coming close to converging with the actual values. The EOL estimates for both procedures in Table 2 are close to the actual EOL values. The LSTM prediction for autoregression is less accurate than GPR as it reduces the influence of the characterisation disturbances (Ma et al., 2019). The LSTM free simulation estimates are significantly more accurate - approximately 6 hours away from the actual value. Thus, the long term trend recognition comes at a slight cost of short term accuracies. In an online setting where the model would be expected to operate in autoregression and free simulation modes, LSTMs would be the more attractive option in terms of accuracy when compared to GPR. However, to suitably apply corrective actions to improve the health and/or lifetime of the fuel cell, it is beneficial to have a probabilistic measure associated with the predictions (Jouin et al., 2014). To combine the probabilistic measures provided by GPRs and the long term trend learning provided by LSTMs, GP-LSTMs are employed.

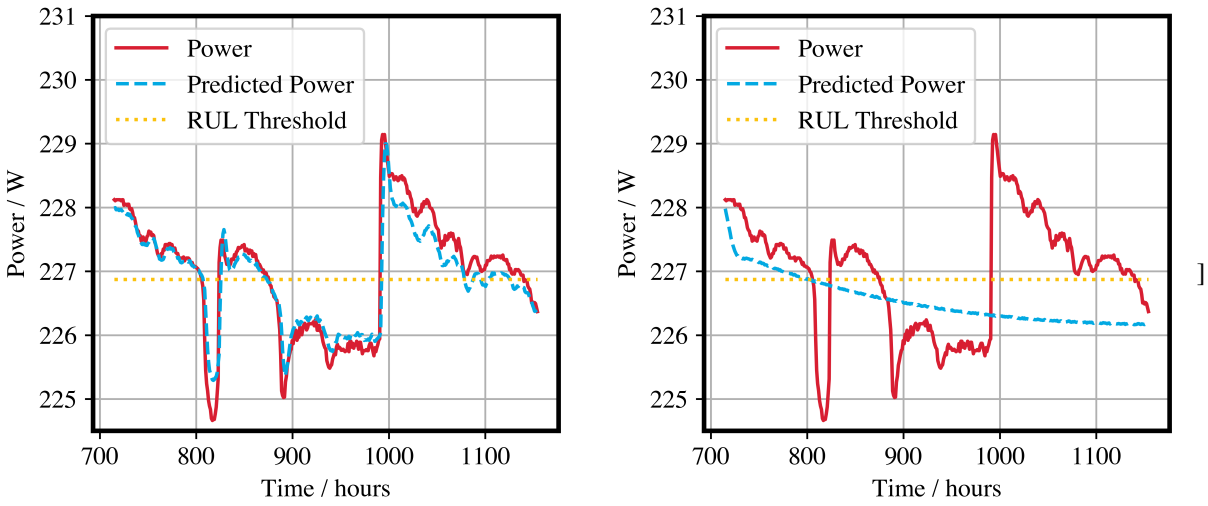
4.3. GP-LSTM

To obtain predictions with an associated probability, GP-LSTMs are applied and the corresponding results are shown in Figure 6. The results were obtained using a kernel comprising of an LSTM layer of 50 units, with $\tanh$ as the activation function and a dropout and recurrent dropout of 0.001 and 0.025 respectively. The number of units and the activation function were kept the same as the previous LSTM model (Al-Shedivat et al., 2017). However, since Gaussian Processes use a separate hyperparameter for uncertainty - likelihood (Murphy, 2014), the dropout values were significantly lowered to provide regularisation but not significantly impact the uncertainty measurements. The specific values were chosen based on their fit over the training measurements. The autoregression results in Figure 6a are very accurate prior to the substantial spike around 1000 hours. After that, the predictions under-predict but appear to converge, albeit slowly. This implies that the effect of the characterisation disturbances is being mitigated, in a manner similar to LSTMs, and similar to the LSTM-ARIMA fusion model suggested by (Ma et al., 2019). Similar to LSTMs, the mitigation can be attributed to the long term trends coming into play and influencing short term results. The uncertainty in the predictions seems to increase with time and around strong transitions, such as the one close to 1000 hours. The RMSE and MAPE estimates for GP-LSTM in Table 1 are slightly lower than the other methods.

However, in free simulation, the GP-LSTM model predicts the general trend before the sharp rise at 1000 hours much better than the other models - as is observable in Figure 6b. The uncertainty, as expected, increases over time in the free simulation mode. Even though the model does not predict the rise after 1000 hours, it is much better at following the trend prior to that. Furthermore, the uncertainty sharply increases after 1000 hours, reflecting a severe lack of confidence in the values predicted. This, in an online setting, would indicate the need to incorporate more measurements before implementing any decisions or corrective actions. The EOL estimates in Table 2 are also quite accurate, especially the free simulation result which is off by approximately 15 hours. Certain hyperparameters can be adjusted to improve the free simulation performance or the autoregression performance, but usually at the cost of the other.

4.3.1. Applying autoregression and free simulation

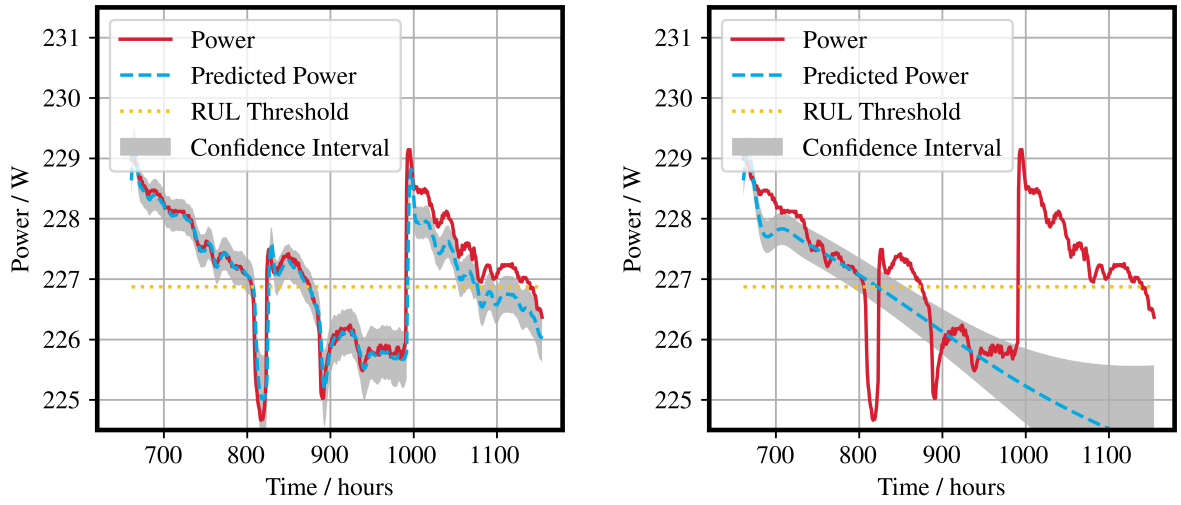
Free simulation results can be improved by increasing the dropout and recurrent dropout. This would help the model generalise better but at the cost of autoregression accuracies. Correspondingly, the dropouts can be decreased slightly for improving autoregression performance. However, the two procedures can be combined in an online prognostic setting for EOL estimations by applying them to different parts of the



(a) LSTM Autoregression

(b) LSTM Free Simulation

Figure 5. LSTM using a sliding window



(a) GP-LSTM Autoregression

(b) GP-LSTM Free Simulation

Figure 6. GP-LSTM using a sliding window

Table 2. Comparing predicted EOL estimates against the actual EOL of 805.57 hours

Model	Autoregression	Free Simulation	Autoregression	Free Simulation
	Predicted EOL		%Er	
GPR	806.36	742.97	0.10	7.77
LSTM	807.07	799.79	0.19	0.72
GP-LSTM	807.20	821.71	0.20	2.00

data as shown in Figures 7 and 8. In Figure 7 it is assumed that in an online setting, the model has only been trained until 350 hours. However, power data between 350 and 500 hours becomes progressively available so autoregression is applied. At each step within this range, the predictions are made based on the true power values for all preceding times. At 500 hours, the model aims to predict ahead without progressively available data and thus, free simulation is applied. The autoregression results fit the available data quite well barring the slight under-prediction around 370 hours which has a sharp transition. The free simulation results are also quite accurate as they predict the general long term trend quite well. The predictions deviate from the actual results around 1000 hours where there is a sharp improvement in performance due to the characterisation test.

Similarly, Figure 8 shows the results with a larger training set and similarly sized autoregression set. Like before, the autoregression results fit the actual values. However, in free simulation, the model's long term trend realisation overwhelms the predictions and the model under predicts. Therefore, in the future, the model will need to be adapted to make it more sensitive to short term trends. (Al-Shedivat et al., 2017) recommend the use of structured prediction algorithms such as DAGGER (Ross, Gordon, & Bagnell, 2011) to improve the performance of the model in free simulation.

4.3.2. Applying GP-LSTM to Dynamic Data Set

To further prove the efficacy of the model, it is applied to the dynamic data set from FC2. The model, without modifying the hyperparameters, is retrained using data until 500 hours and the results are shown in Figure 9. The results are definitely not accurate, especially the free simulation results. The high deviation of the predictions from the actual measurements can be attributed to the lack of training variables that inhibit it from learning the variations caused by the dynamic current load, a factor that was not significant for FC1 since FC1 was subjected to stationary current loads. To account for this, current is included as an input variable for the model. The number of units in the LSTM layer is increased to 60 to account for the additional inputs but the other hyperparameters are kept the same. Figure 10 shows the autoregression and free simulation results. The model in auto-regression, shown in Figure 10a, generally under-predicts power, especially after spikes. However, its learned long term tendencies still allow it to converge the actual results at the end. The free simulation results in Figure 10b also under-predict generally but also converge with the actual values towards the end. Unlike the results for FC1, the results for FC2 have extremely high uncertainties in free simulation, resulting from a much more dynamic input set. In order to improve its uncertainty for predictions, more training data would be required. The free simulation performance could also be improved by employing structured prediction methods like DAGGER (Ross et al., 2011; Al-Shedivat et al., 2017).

5. CONCLUSION

This study explores the use of GPR and LSTM for prognostics and assesses them in terms of their suitability for online RUL estimation by applying them in autoregression and free simulation contexts. While both methods display highly accurate results for autoregression, GPR fails to appropriately capture the long term tendencies of the data. On the other hand, LSTM is a lot more effective at learning long term trends but does not provide an uncertainty - a characteristic beneficial in an online prognostics setting, especially if corrective action is to be applied. In order to combine the strengths of both methods, the use of GP-LSTM is explored (Al-Shedivat et al., 2017).

GP-LSTM is able to determine long term tendencies and provides an uncertainty measure. It provides autoregression and free simulation results on par with LSTM when applied to the FC1 data set - their performances are very similar for practical purposes. It also reduces the influence of disturbances and favours the long term trends. This study further proposes the use of GP-LSTM in both autoregression and free simulation contexts simultaneously. Here, autoregression helps update the estimates constantly and is used for short term trends while free simulation provides a window into the long term trends. The efficacy of the method is further demonstrated by applying it on a dynamic data set.

To properly establish the efficacy of the model in an online context, the timing associated with the model training and predictions will need to be considered in more detail. Moving forward, structured prediction methods like DAGGER can be incorporated to make the free simulation results more sensitive to short term trends. Since PEMFC performance can also fluctuate with ambient conditions and operating parameters, the methodology can be expanded to account for these factors. The method can also be explored in even more dynamic

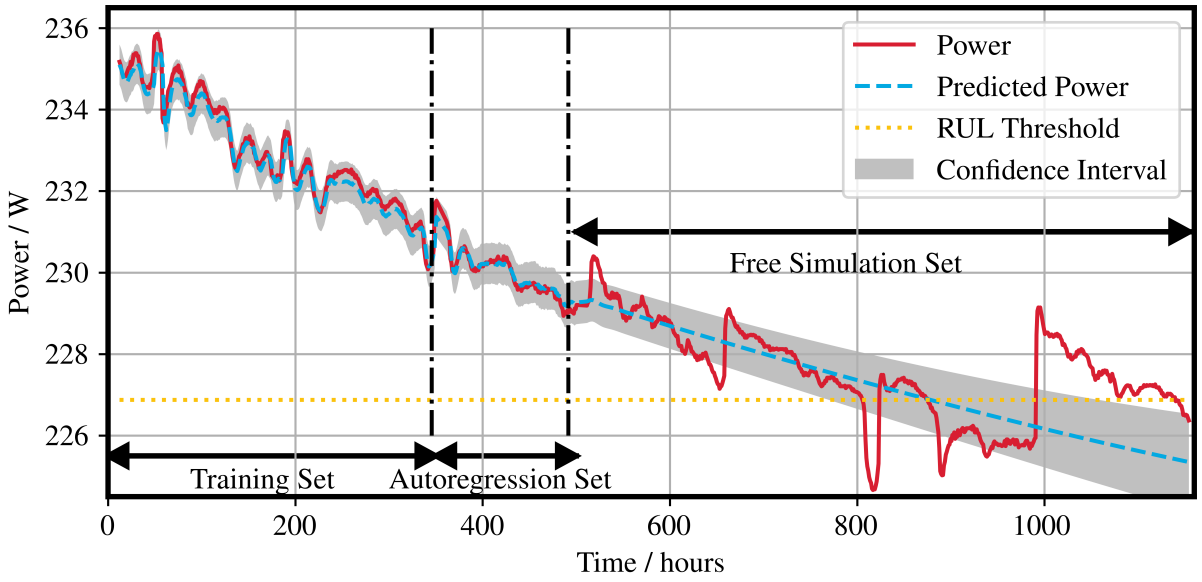


Figure 7. Applying GP-LSTM in phases with autoregression starting at 350 hours and free simulation starting at 500 hours

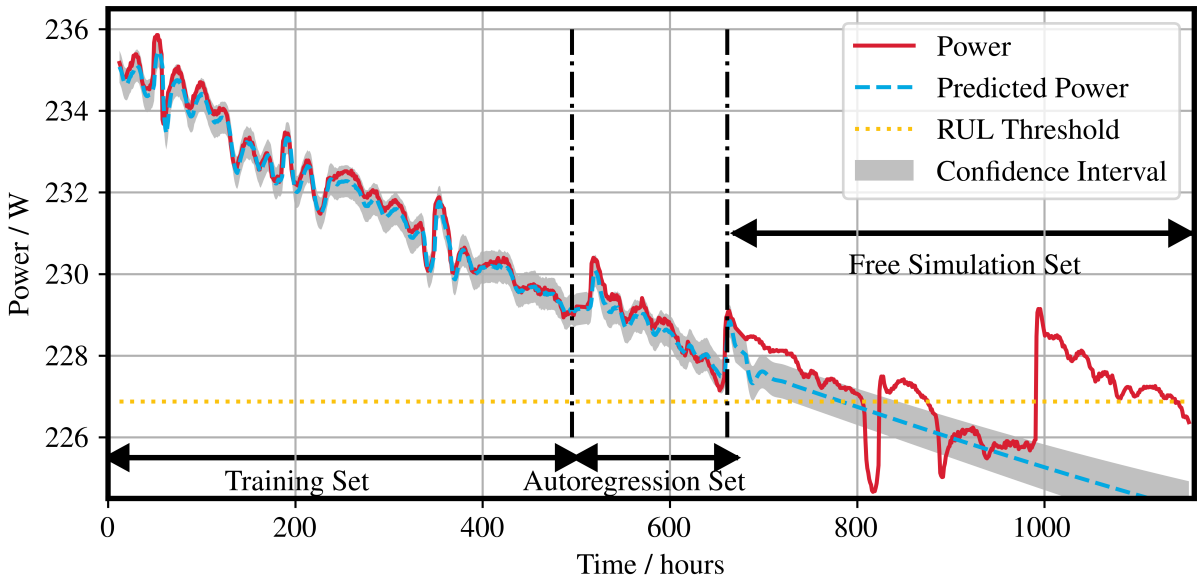
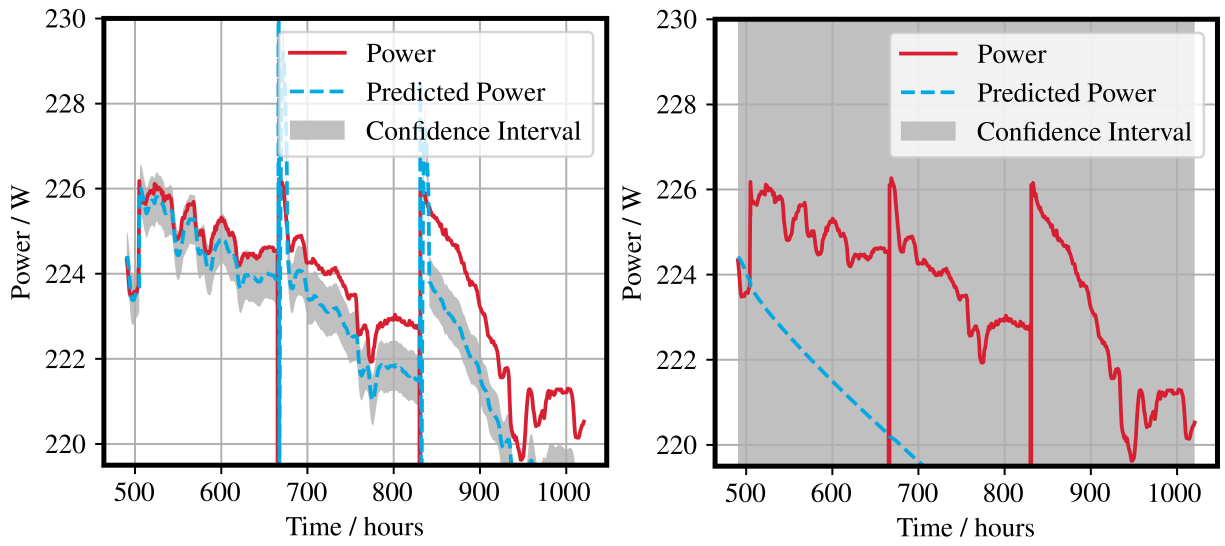


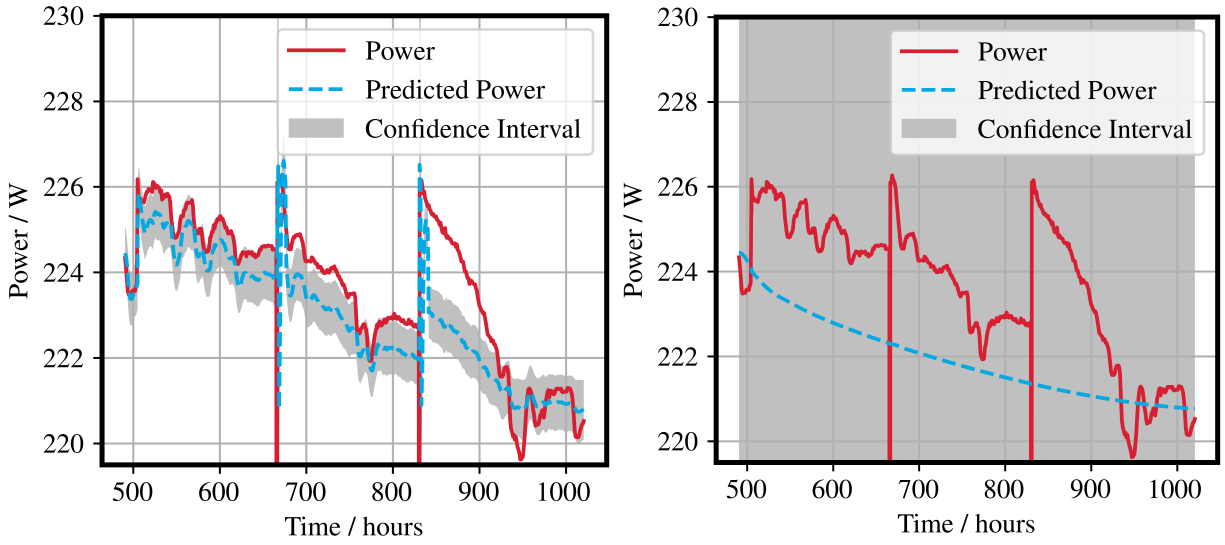
Figure 8. Applying GP-LSTM in phases with autoregression starting at 500 hours and free simulation starting at 670 hours



(a) GP-LSTM Autoregression

(b) GP-LSTM Free Simulation

Figure 9. GP-LSTM for the dynamic data set without using current as an input variable



(a) GP-LSTM Autoregression

(b) GP-LSTM Free Simulation

Figure 10. GP-LSTM for the dynamic data set with current as an additional input variable

settings, such as automotive profile proposed by (Gonzalez et al., 2019).

REFERENCES

Al-Shedivat, M., Wilson, A. G., Saatchi, Y., Hu, Z., & Xing, E. P. (2017). Learning scalable deep kernels with recurrent structure. *Journal of Machine Learning Research*, *18*, 1–37.

Boškoski, P., Debenjak, A., & Boshkoska, B. M. (2017). *Fast electrochemical impedance spectroscopy : as a statistical condition monitoring tool*. Cham, Switzerland: Springer International Publishing.

Bressel, M., Hilairet, M., Hissel, D., & Ould Bouamama, B. (2016). Extended Kalman Filter for prognostic of Proton Exchange Membrane Fuel Cell. *Applied Energy*, *164*, 220–227.

Chen, K., Laghrouche, S., & Djerdir, A. (2019). Degradation model of proton exchange membrane fuel cell based on a novel hybrid method. *Applied Energy*, *252*(June), 113439.

Cheng, Y., Zerhouni, N., & Lu, C. (2018). A hybrid remaining useful life prognostic method for proton exchange membrane fuel cell. *International Journal of Hydrogen Energy*, *43*(27), 12314–12327.

FCLAB. (2014). *IEEE PHM 2014 Data Challenge* (Tech. Rep.). FCLAB Research Federation.

Gong, A., Palmer, J. L., Brian, G., Harvey, J. R., & Verstraete, D. (2016). Performance of a hybrid, fuel-cell-based power system during simulated small unmanned aircraft missions. *Int. J. Hydrogen Energy*, *41*(26), 11418–11426.

Gong, A., & Verstraete, D. (2017). Fuel cell propulsion in small fixed-wing unmanned aerial vehicles: Current status and research needs. *Int. J. Hydrogen Energy*, *42*(33), 21311–21333.

Gonzalez, E. L., Cuesta, J. S., Fernandez, F. J. V., Llerena, F. I., Carlini, M. R., Bordons, C., ... Elfes, A. (2019). ScienceDirect Experimental evaluation of a passive fuel cell / battery hybrid power system for an unmanned ground vehicle pez Gonz a. *International Journal of Hydrogen Energy*, *44*(25), 12772–12782.

Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow : concepts, tools, and techniques to build intelligent systems* (Second edition. ed.). Sebastopol, CA: O'Reilly.

Javed, K., Gouriveau, R., Zerhouni, N., & Hissel, D. (2016a). PEM fuel cell prognostics under variable load: A data-driven ensemble with new incremental learning. *International Conference on Control, Decision and Information Technologies, CoDIT 2016*, 252–257.

Javed, K., Gouriveau, R., Zerhouni, N., & Hissel, D. (2016b). Prognostics of Proton Exchange Membrane Fuel Cells stack using an ensemble of constraints based connectionist networks. *Journal of Power Sources*, *324*, 745–757.

Jha, M. S., Bressel, M., Ould-Bouamama, B., & Dauphin-Tanguy, G. (2016). Particle filter based hybrid prognostics of proton exchange membrane fuel cell in bond graph framework. *Computers and Chemical Engineering*, *95*, 216–230.

Jouin, M., Gouriveau, R., Hissel, D., Péra, M. C., & Zerhouni, N. (2013). Prognostics and Health Management of PEMFC - State of the art and remaining challenges. *International Journal of Hydrogen Energy*, *38*(35), 15307–15317. doi: 10.1016/j.ijhydene.2013.09.051

Jouin, M., Gouriveau, R., Hissel, D., Péra, M. C., & Zerhouni, N. (2014). Prognostics of PEM fuel cell in a particle filtering framework. *International Journal of Hydrogen Energy*, *39*(1), 481–494.

Jouin, M., Gouriveau, R., Hissel, D., Péra, M. C., & Zerhouni, N. (2015). PEMFC aging modeling for prognostics and health assessment. *IFAC-PapersOnLine*, *28*(21), 790–795.

Kimotho, J. K., Meyer, T., & Sextro, W. (2015). PEM fuel cell prognostics using particle filter with model parameter adaptation. *2014 International Conference on Prognostics and Health Management, PHM 2014*, 1–6.

Larminie, J., & Dicks, A. (2001). *Fuel Cell Systems Explained* (Vol. 93).

Liu, H., Chen, J., Hissel, D., Lu, J., Hou, M., & Shao, Z. (2020). Prognostics methods and degradation indexes of proton exchange membrane fuel cells: A review. *Renewable and Sustainable Energy Reviews*, *123*(February), 109721.

Liu, H., Chen, J., Hissel, D., & Su, H. (2019). Short-Term Prognostics of PEM Fuel Cells: A Comparative and Improvement Study. *IEEE Transactions on Industrial Electronics*, *66*(8), 6077–6086.

Liu, H., Chen, J., Hou, M., Shao, Z., & Su, H. (2017). Data-based short-term prognostics for proton exchange membrane fuel cells. *International Journal of Hydrogen Energy*, *42*(32), 20791–20808.

Liu, H., Ong, Y.-S., Shen, X., & Cai, J. (2020). When Gaussian Process Meets Big Data: A Review of Scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems*, *PP*, 1–19.

Liu, J., Li, Q., Chen, W., Yan, Y., Qiu, Y., & Cao, T. (2019). Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks. *International Journal of Hydrogen Energy*, *4*, 5470–5480.

Ma, R., Breaz, E., Liu, C., Bai, H., Briois, P., & Gao, F. (2018). Data-driven Prognostics for PEM Fuel Cell Degradation by Long Short-term Memory Network. *2018 IEEE Transportation and Electrification Conference and Expo, ITEC 2018*, 1094–1099.

Ma, R., Li, Z., Breaz, E., Liu, C., Bai, H., Briois, P., & Gao, F. (2019). Data-fusion prognostics of proton exchange membrane fuel cell degradation. *IEEE Transactions on Industry Applications*, *55*(4), 4321–4331.

Ma, R., Yang, T., Breaz, E., Li, Z., Briois, P., & Gao, F. (2018). Data-driven proton exchange membrane fuel cell degradation predication through deep learning method. *Applied Energy*, *231*(July), 102–115.

Murphy, K. (2014). *Machine Learning, a Probabilistic Perspective* (Vol. 27) (No. 2).

Qin, S., Zhu, J., Qin, J., & Wang, W. (2019). Recurrent Attentive Neural Process for Sequential Data. , 1–12.

Richardson, R. R., Osborne, M. A., & Howey, D. A. (2019). Battery health prediction under generalized conditions using a Gaussian process transition model. *Journal of Energy Storage*, *23*(April), 320–328.

Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., & Aigrain, S. (2013). Gaussian processes for time-series modelling.

Ross, S., Gordon, G. J., & Bagnell, J. A. (2011). A reduction of imitation learning and structured prediction to no-regret online learning. *Journal of Machine Learning Research*, *15*, 627–635.

Sharaf, O. Z., & Orhan, M. F. (2014). An overview of fuel cell technology: Fundamentals and applications. *Renewable Sustainable Energy Rev.*, *32*, 810–853.

Sutharssan, T., Montalvao, D., Chen, Y. K., Wang, W.-C., Pisac, C., & Elemara, H. (2017). A review on prognostics and health monitoring of proton exchange membrane fuel cell. *Renewable Sustainable Energy Rev.*, *75*(November 2015), 440–450.

Verstraete, D., Gong, A., Lu, D. D. C., & Palmer, J. L. (2015). Experimental investigation of the role of the battery in the AeroStack hybrid, fuel-cell-based propulsion system for small unmanned aircraft systems. *Int. J. Hydrogen Energy*, *40*(3), 1598–1606.

Verstraete, D., Lehmkuehler, K., Gong, A., Harvey, J. R., Brian, G., & Palmer, J. (2014). Characterisation of a hybrid, fuel-cell-based propulsion system for small unmanned aircraft. *J. Power Sources*, *250*, 204–211.

Wang, F. K., Cheng, X. B., & Hsiao, K. C. (2020). Stacked long short-term memory model for proton exchange membrane fuel cell systems degradation. *Journal of Power Sources*, *448*(December 2019), 227591.

Wu, Y., Breaz, E., Gao, F., Paire, D., & Miraoui, A. (2016). Nonlinear performance degradation prediction of proton exchange membrane fuel cells using relevance vector machine. *IEEE Transactions on Energy Conversion*, *31*(4), 1570–1582.

Yang, C., Li, Z., Liang, B., Lu, W., Wang, X., & Liu, H. (2017). A particle filter and long short term memory fusion algorithm for failure prognostic of proton exchange membrane fuel cells. *Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017*, 5646–5651.

Zhou, D., Gao, F., Breaz, E., Ravey, A., & Miraoui, A. (2017). Degradation prediction of PEM fuel cell using a moving window based hybrid prognostic approach. *Energy*, *138*, 1175–1186.

Zhu, L., & Chen, J. (2018). Prognostics of PEM fuel cells based on Gaussian process state space models. *Energy*, *149*, 63–73.

Zhu, S., Yuan, X., Xu, Z., Luo, X., & Zhang, H. (2019). Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast. *Energy Conversion and Management*, *198*(June), 111772.