# A Real-time Data-driven Method for Battery Health Prognostics in Electric Vehicle Use

Anthony Barré[1 2], Frédéric Suard[1], Mathias Gérard[2], and Delphine Riu[3]

[1] *CEA, LIST, Information Models and Machine Learning, 91191 Gif sur Yvette CEDEX, France*
*anthony.barre@cea.fr*
*frederic.suard@cea.fr*

[2] *CEA, LITEN, 17 rue des martyrs, 38054 Grenoble CEDEX 9, France*
*mathias.gerard@cea.fr*

[3] *G2Elab, UMR 5269, 38402 Saint Martin d'Hères, France*
*delphine.riu@g2elab.grenoble-inp.fr*

## ABSTRACT

Online prognostics of the battery capacity is a major challenge as ageing process is a complex phenomenon, hardly directly measurable. This paper offers a new methodology for real-time estimating of the global battery performances for Electric Vehicle (EV) use. The presented data-driven framework build a model based on the modifications in battery signals behavior, according to the performance level. A first pattern extraction step consists in the selection of battery signals corresponding to specific acceleration profiles in real uses, allowing to highlight the battery behavior. These extracted voltage and current patterns are then considered to determine the battery behavior for each State of Health (SOH) feature. Studied patterns are compared using signal processing techniques, allowing the estimation of the battery performance, through statistical learning methods. The application of signal processing and Relevance Vector Machines (RVM) model with multiple kernels, provides a powerful tool to diagnose battery health online, only based on real signals. Furthermore, this methodology also allows the prediction of battery Remaining Useful Life (RUL) during real use. The proposed algorithm is validated using datasets from real EV uses. Presented diagnostics results on real data demonstrate the good accuracy of this new framework for battery SOH prognostics in real-time constraints, with uncontrolled conditions.



## 1. INTRODUCTION

Lithium-ion (Li-ion) batteries are becoming the battery of choice in Electric Vehicles (EV) utilization. However, battery health and lifetime remain a major drawback to the use of Li-ion batteries in stringent life requirements. In EV context, accurate battery health assessment is primordial to improve the users confidence in the battery range. Indeed, it is one of the biggest obstacles to widespread acceptance of EVs. Market experts evaluated the effects of low range resources of EVs, as a significant feature for users' purchases intentions (Peters & Dütschke, 2014).

The field of prognostics and health management offers different approaches for estimating battery age level and remaining lifetime (Saha & Goebel, 2008). There are many data-driven methodologies that focus on the battery State of Health (SOH) estimation (Barré et al., 2013). However, most of these data-driven approaches perform well on their training data only, under specific operational experiments, inducing robustness and generalization mistakes. In real life, external conditions cannot be controlled and these learned models are subject to misestimations. Thus, an accurate way of estimating battery capacity in real-time based on real EV uses data-driven algorithm still requires investigations (Barré, Suard, Gérard, Montaru, & Riu, 2014).

In this work, we propose an alternative approach by only using data-driven methodology developed from a set of real EV uses. Such a methodology requires large amounts of training data in the development phase. In the EV context this training data requirement is very restrictive and costly. To face this problem, we investigate whether it is possible to extract relevant features from current and voltage signals collected during real EV uses, under non controlled conditions. A key

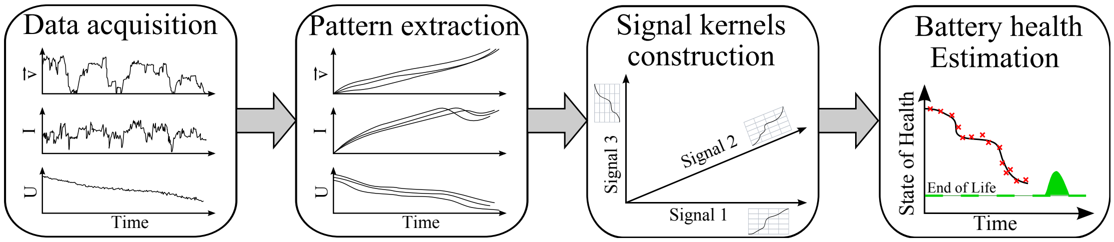



Figure 1. Battery signal analysis framework consisting of the data aquisition, the extraction of specific patterns, the formation of new feature space and the estimation of the battery health level along with its remaining useful life

issue explored by this paper is how battery capacity can be estimated during real EV uses, without specific requirement, based only on real use data.

Section 2 presents the global theoretical framework and details the methods used for the SOH estimation and RUL prognostic in real time. Then, Section 3 details the obtained results for SOH estimation and RUL prognosis in real EV uses context. Finally, Section 4 presents the main conclusions and discussions of this research.

## 2. METHODOLOGY

In this section, we present our approach to predict the battery State of Health (SOH) and its corresponding Remaining Useful Life (RUL). As a first step, signal patterns were estimated from the acquired data, in order to observe the battery behavior modification. To quantify these modifications Dynamic Time Warping (DTW) is introduced. Then, a multiple kernel Relevance Vector Machines (RVM) model is learned to estimate the battery SOH in real time. Based on the SOH estimations, the RUL is predicted with a bootstrap approach. The global framework is illustrated in Figure 1.

### 2.1. Patterns extraction

The proposed methodology is based on the assumption of battery behavior modification along the battery life. Time series signal can be used to diagnose health by analyzing battery behavior. Thus, for a similar battery request, it is possible to detect battery ageing effects based on signal shapes such as current and voltage.

To observe a signal behavior modification, it is necessary to compare battery signals under comparable uses. For example, during an identical speed profile criterion, the battery voltage does not react the same way depending on its health level. Thus, the common reference here is the speed signal. In the following study we consider maximal accelerations from 10 to 60 km/h in less than 12.5 seconds as a reference criterion. This choice allows to extract patterns with a relatively large length permitting to detect battery behavior. These extracted training data are consequently issued from real EV uses, providing a large amount of data under uncontrolled conditions.

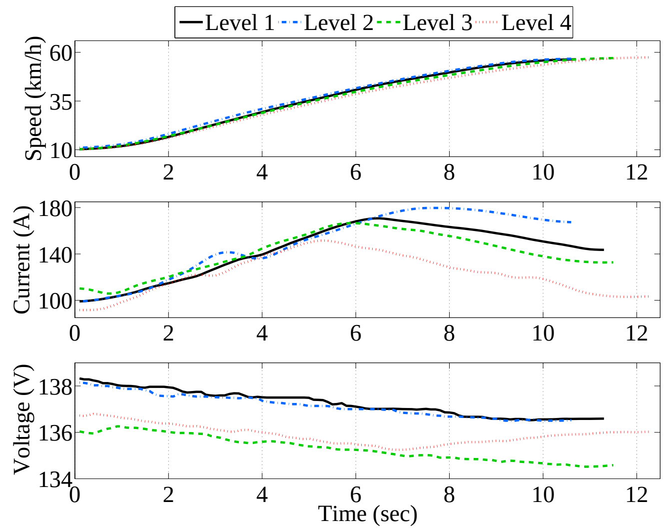



Figure 2. Average signals profile corresponding to a maximal acceleration under four different health levels, based on 10-60 km/h accelerations - Level 1 (SOH = 100 - 98%, T $\simeq$ 24°C) - Level 2 (SOH = 98 - 96%, T $\simeq$ 22°C) - Level 3 (SOH = 96 - 92%, T $\simeq$ 12°C) - Level 4 (SOH = 92 - 87%, T $\simeq$ 30°C)

To compare the pattern behaviors under different health levels we proceed to an average shape of the patterns for different battery health classes. Thus, the average shapes of the extracted signals, under four different health classes are presented in Figure 2. These classes represent four battery health levels, sorted from the least period "Level 1" to the most aged battery level "Level 4". The SOH level used to build these groups, is here extrapolated from several complete characterizations permitting to obtain SOH reference values. The average temperature values of each class are various, going from 12°C for Level 3 to 30°C for Level 4. It is important to note that these temperature conditions can induce potential pattern behavior modification.

Figure 2 illustrates the variations of signals behavior for different battery health levels. The extracted speed profiles are really close to each other, forming good comparative samples. However, in real-life context it is impossible to obtain exactly twice the same speed profile, implying a slight diversity among the extracted speed signals.

On the contrary, the corresponding extracted current and voltage patterns have various shapes. For example, the current pattern of Level 4 (SOH between 92 and 87 %) is clearly below the other ones, and its corresponding voltage pattern increases faster than for the other health classes. This can be explained by a difference of battery reaction for a same power demand at different health levels. These behavior modifications demonstrate the alteration of the battery reaction in correlation with health degradation.

The objective of the following study is to use these battery behavior modifications to estimate its health level, only based on the extracted patterns.

### 2.2. Dynamic Time Warping

In order to compare the signals pattern and quantify their similarities, we have to consider a metric adapted to this problem. Thus, beyond usual measures, the current state-of-the-art of shape similarity quantification is the Dynamic Time Warping (DTW). It permits to compare asynchronous signals of different lengths. The primary goal of DTW is to compare sequences respecting their shapes by finding an optimal alignment function stretching them. Since its introduction in the 70s, DTW has commonly been used in signals similarity problems in many fields : speech processing, signals recognition, data mining and imaging (Aach & Church, 2001; Bar-Joseph, Gerber, Gifford, Jaakkola, & Simon, 2002; Petitjean, Kurtz, Passat, & Ganarski, 2012).

This method is based on the Levenshtein distance (Sakoe & Chiba, 1971) and finds the optimal path between two sequences, considering temporal distortion. This optimal path produces an alignment function, along with a shape-based similarity measure. Formally, we have two sequences $X := (x_1, .., x_N)$ of length $N \in \mathbb{N}$ and $Y := (y_1, ..., y_M)$ of length $M \in \mathbb{N}$. In the following we fix a feature space denoted by $\mathcal{F}$. To compare two different features $x, y \in \mathcal{F}$, one needs a local cost measure, defined by a function $c$ :

$$c : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_{\geq 0} \tag{1}$$

Typically, the cost $c(x, y)$ is low if $x$ and $y$ are similar to each other, otherwise $c(x, y)$ is high. Evaluating the local cost measure $c(x, y)$ for each pair of elements of the sequences $X$ and $Y$, one obtains the cost matrix $C \in \mathbb{R}^{N \times M}$ defines by $C(i, j) = c(x_i, y_j)$. The goal is to find the alignment between $X$ and $Y$ minimizing the overall cost. A warping path is a sequence $p = (p_1, ..., p_L)$ with $p_l = (n_l, m_l) \in [1 : N] \times [1 : M]$, $\forall l \in [1 : L]$, satisfying the conditions :

$$\begin{cases} p_1 = (1,1) \text{ and } p_L = (N, M) \\ n_1 \leq ... \leq n_L \text{ and } m_1 \leq ... \leq m_L \\ p_{l+1} - p_l \in \{(1,0), (0,1), (1,1)\},\ \forall l \in [1 : L-1] \end{cases} \tag{2}$$

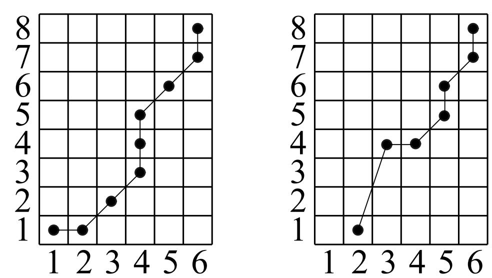
A warping path $p = (p_1, ..., p_L)$ defines an alignment between two sequences $X$ and $Y$ by assigning the element $x_{n_l}$ of $X$ to the element $y_{m_l}$ of $Y$. The alignment conditions imply that the first elements of $X$ and $Y$ as well as their last elements are aligned to each other. The total cost $c_p(X, Y)$ of a warping path $p$ between $X$ and $Y$ with respect to the local cost measure $c$ is defined as :

$$c_p(X, Y) = \sum_{l=1}^{L} c(x_{n_l}, y_{m_l}) \tag{3}$$

Furthermore, an optimal warping path between $X$ and $Y$ is a warping path $p^*$ minimizing total cost among all possible warping paths. The DTW distance $d_{DTW}(X, Y)$ between $X$ and $Y$ is then defined as the total cost of the optimal warping path $p^*$ :

$$d_{DTW}(X, Y) = c_{p^*}(X, Y) = \min \{c_p(X, Y) | \, \forall p\} \tag{4}$$

The local cost measure $c$ is defined as the distance between elements of sequences, e.g., the Euclidean distance. An example of DTW warping paths is given in Figure 3.



(a) (b)

Figure 3. Illustration of paths of index pairs for a sequence $X$ of length $N = 6$ and a sequence $Y$ of length $M = 8$ (a) Admissible warping path (b) Example of a non admissible warping path due to boundary conditions and step size conditions

This DTW distance permits the comparison and the quantification of different signals shape. It is particularly adapted to battery signals evolution. Therefore, this distance measures the difference between each extracted pattern.

### 2.3. Relevance Vector Machines

The Relevance Vector Machines (RVM), initially introduced by (Tipping, 2001), is based on a Bayesian formulation of a linear model with an appropriate prior that results in a sparse representation. Given the set of training patterns $\{t_i|i = 1, ..., N\}$ along with their corresponding health level $\{h_i|i = 1, ..., N\}$, assume that $h_i = f(t_i) + \epsilon_i$, where $\epsilon_i$ are assumed to be independent samples from a Gaussian noise process with zero mean and $\sigma^2$ variance, i.e. $\sigma_i \sim \mathcal{N}(0, \sigma^2)$, $\forall i$. The aim is to learn a dependency model of the targets on the inputs to make accurate predictions of $h$ for unseen values of $t$. Typically, predictions are based on some function $f(t)$ defined over the input space, and learning is the process of inferring the parameters of this function. This function takes the form :

$$f(t) = \sum_{i=1}^{M} w_i K(t, t_i) + w_0 \tag{5}$$

where $f(t)$ is the function output, $K(t, t_i)$ is a kernel function and $w = [w_1, ..., w_N]^T$ are the weights.

Therefore, the likelihood of dataset can be written as :

$$p(h|w, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\{-\frac{1}{2\sigma^2}||h - \phi w||^2\} \tag{6}$$

where $\phi = [\phi(t_1), ..., \phi(t_N)]^T$,
and $\phi(t_N) = [1, K(t_i, t_1), K(t_i, t_2), ..., K(t_i, t_N)]^T$

When attempting to learn the relationship between $t$ and $h$, we wish to constrain complexity and hence the growth of the weights $w$ by defining an explicit prior probability distribution on $w$. Our preference for smoother and therefore less complex functions is encoded by using a zero-mean Gaussian prior over $w$. This gives us :

$$p(w|\alpha) = \sum_{i=1}^{N} N(0, \alpha_i^{-1}) \tag{7}$$

where we have used $\alpha_i$ to describe the inverse variance of each $w_i$. This means that there is a hyperparameter $\alpha_i$ associated with each weight, modifying the strength of the prior thereon. To complete the specification of this hierarchical prior, we must define hyperpriors over $\alpha$; as well as over the noise variance $\sigma^2$.

Having defined the prior, Bayesian inference proceeds by computing the posterior over all unknows given the data from Bayes' rule, i.e. :

$$p(w, \alpha, \sigma^2|h) = \frac{p(h|w, \alpha, \sigma^2)p(w, \alpha, \sigma^2)}{p(h)} \tag{8}$$

Assuming that the new test target is $h_*$, and the new test input $t_*$ are used to make predictions. The predictions are made according to :

$$p(h_*|h) = \int p(h_*|w, \alpha, \sigma^2)p(w, \alpha, \sigma^2|h)dwd\alpha d\sigma^2 \tag{9}$$

We can decompose the posterior $p(w, \alpha, \sigma^2|h)$ as :

$$p(w, \alpha, \sigma^2|h) = p(w|h, \alpha, \sigma^2)p(\alpha, \sigma^2|h) \tag{10}$$

And so, the posterior distribution over the weights is :

$$p(w|h, \alpha, \sigma^2) = \frac{p(h|w, \alpha, \sigma^2)p(w|\alpha)}{p(h|\alpha, \sigma^2)} \sim \mathcal{N}(w|\mu, \Sigma) \tag{11}$$

where the posterior covariance and mean are respectively :

$$\Sigma = (\sigma^{-2}\phi^T\phi + A)^{-1} \tag{12}$$

$$\mu = \sigma^{-2}\Sigma\phi^T h \tag{13}$$

with $A = diag(\alpha_0, ..., \alpha_N)$. Note that $\sigma^2$ is also treated as hyperparameter, which can be estimated from the data.

Therefore, machine learning becomes a search for the most probable hyperparameters posterior $\alpha_{MP}$ and $\sigma^2_{MP}$. Predictions for a new input data $t_*$ are made according to the integration of weights to obtain the marginal likelihood for the hyperparameters :

$$p(h|\alpha_{MP}, \sigma^2_{MP}) = \int p(h_*|w, \sigma^2_{MP})p(w|\alpha_{MP}, \sigma^2_{MP})dw$$

$$p(h|\alpha_{MP}, \sigma^2_{MP}) = \mathcal{N}(h_*|t_*, \sigma^2_*) \tag{14}$$

with :

$$h_* = \mu^T\phi(t_*) \tag{15}$$

$$\sigma^2_* = \sigma^2_{MP} + \phi(t_*)^T\Sigma\phi(t_*) \tag{16}$$

In order to employ the DTW measure in the RVM process, we have to use a kernel function $K$ considering the DTW measure. Several attempts were made to derive kernels based on the DTW distance (Lei & Sun, 2007). We consider in this paper the Gaussian Dynamic Time Warping (GDTW) kernel (Bahlmann, Haasdonk, & Burkhardt, 2002), with a parameter $\gamma$, defined as :

$$K_{GDTW}(t, t_i) = exp(-\gamma d_{DTW}(t, t_i)) \tag{17}$$

### 2.4. Extension of RVM to Multiple Kernel

The use of different kernels in a same process allows the combination of different characteristics. In the case of complex phenomena, a multiple kernel approach can be useful as each used kernel is subject to extract a different characteristic, obtained with different kernel formations or parameters (Suard & Mercier, 2009). In order to assign specific kernel for each patterns, the prediction function is written :

$$f(t) = \sum_{i=1}^{N}\sum_{j=1}^{k} w_{i,j} \cdot K_j(t, t_i) + w_0 \tag{18}$$

One possible way to write this function is to define a kernel basis. This definition decomposes the kernel $K$ into different blocks. The multiple kernel, for $k$ kernels, is then composed like a kernel basis :

$$K = [1\ K_1\ K_2\ ...\ K_k] \tag{19}$$

If we consider that all columns are independent, we can finally write the prediction function with :

$$f(t) = \sum_{i=1}^{N*k} w_i \cdot K_i(t) + w_0 \tag{20}$$

Thus, this formulation shows that we can extend RVM to multiple kernel with a kernel basis approach.

## 2.5. RUL prediction

The fitted model is used to estimate battery health at different times. Based on these SOH estimations $\{h_{i*} | i = 1, ..., N\}$, the aim is to predict the battery Remaining Useful Life (RUL). The RUL is defined as the remaining time until the battery reaches an End of Life (EOL) criterion, commonly chosen as $80\%$ SOH level.

Remaining Useful Life (RUL) is derived by projecting out the capacity estimates into the future until expected capacity hits the certain predetermined End of Life (EOL) threshold. As opposed to the SOH estimations, this process does not require to be done in real time as the SOH dynamic is too slow to modify the RUL at every EV use. Thus, at a given time $T$, the proposed methodology considers a polynomial regression to fit all the past SOH estimations $\{h_{i*} | i = 1, ..., N\}$, with $N \leq T$. The polynomial regression finds the coefficients of a polynomial $p$ of degree $d$ that fits $p(T)$ to the estimated battery health level $h_{i*}$ at a time $i$, in a least square sense. The polynomial $p$ of degree $d$ is defined as :

$$p(T) = p_0 + p_1 T + ... + p_d T^d = \sum_{j=0}^{d} p_j T^j \tag{21}$$

Considering this polynomial construction, the aim is to build a RUL probability density function. For this, we use a bootstrap technique to predict the value of the RUL with a statistical sampling.

Thus, we sample past SOH estimations $\{h_{i*} | i = 1, ..., N\}$, with replacement, obtaining bootstrap data $\{h_{i*}^B | i = 1, ..., N\}$. For this bootstrap data we calculate the corresponding polynomial $p$ and then use this polynomial to predict its associated RUL.

Repeating these steps $L$ times, we obtain a family of bootstrap RUL predictions $\{\widehat{RUL}_g^B | g = 1, ..., L\}$. The distribution of the $\widehat{RUL}_g^B$ allows the construction of a predicted RUL probability density function (pdf) at a time T.

Even with few SOH estimations, this bootstrap permits the obtention of RUL pdf. Note that, in battery RUL prediction context, we will consider in the following study a polynomial degree $d = 2$ to fit the SOH dynamics. This choice is a consequence of the slow variations of SOH evolution observed in the literature.

# 3. RESULTS

## 3.1. Model learning

The described framework is applied on battery real dataset. The considered methodology is here tested with patterns extracted with a 10-60 km/h in less than 12.5 seconds as an acceleration criterion. Note that a long acceleration pattern contains more information and less variability than the short ones. However, the longest acceleration profiles require larger datasets to obtain enough patterns for the methodology process.

The data used in this study wascollected from a real and non-controlled EV use, during 460 days, generating 50 000 km. Thereby processing data are representative of a large variety of conditions an EV battery can be faced with. Moreover, using real data ensures compatibility of the developed methodology for embedded uses. The presented results here come from a unique EV battery to illustrate the methodology performances. This experiment also contains several battery characterizations permitting to measure the real SOH of the battery, through a complex specific process done a test bench. Thus, these measured SOH values compose the targeted health levels $h$, and the aim is to produce SOH estimations $\hat{h}$, with the explained methodology.

To illustrate the frequency of pattern extraction, on the studied real data, an acceleration profile corresponding to the defined criterion happens in average every 150 km of EV use. This value is given here as an indication as it is of course highly dependent on the driving style and to driving conditions. This property induces that during a real EV use, the model provides a new SOH estimation on average every 150 km, which represents a good frequency compared to the total battery lifetime estimated to be approximately of 160 000 km. Thus, for a utilization of 15 000 km per year, the algorithm produces two SOH estimations every week.

The extracted voltage and current patterns, as presented in Figure 1, along with capacity references obtained from specific tests are then used. The training data sets are composed of patterns issued from real EV uses and of frequent battery measurements. Thus, each extracted pattern is associated to a battery health level, permitting the training of the RVM algorithm. In this study we use $75\%$ of the data to train the RVM algorithm and the other $25\%$ compose the test data permitting to evaluate the methodology accuracy.

The first step of the presented methodology is to create a RVM model to estimate online the battery health during its real uses. As explained in Section 2.3, the use of several kernels can add information into the model. Thus, in this study we consider a multiple kernel RVM approach, with the association of three different kernels. Two GDTW kernels are respectively built with the extracted current patterns and with the extracted voltage patterns. The other kernel is a Gaussian kernel calculated from the values of the battery temperature measured at each pattern extraction. Battery temperature measures are here introduced into the model construction to avoid adding information about the variable external conditions. Indeed, it is well known that the battery temperature highly influences its reaction and its signals behavior. The SOH targets $h$ are here the extrapolations of the measures obtained with the battery characterizations. Note that this model learning step is computationally constraining, as it requires a lot of DTW calculations. But this model construction is done just once, before real time application context. The complexity of this step is consequently not a drawback to the application in a real EV use.

Therefore, the inputs of the learning RVM model are the kernels corresponding to the current and voltage patterns along with a kernel based on their associated battery temperature measures, the output is an estimation $\hat{h}$ of the battery SOH level.

### 3.2. Results

Based on the learned RVM model, the methodology allows the SOH diagnosis whenever a specific 10-60 km/h acceleration in less than 12.5 seconds, is detected during EV use. Thus, the embedded trained algorithm produces a new SOH estimation at each acceleration corresponding to our criteria defined in Section 2.1. The pattern extraction step is done in real time, as it only requires a criteria comparison step. Once a speed pattern is detected as satisfying, the extraction criteria, the corresponding voltage and current patterns are then extracted, along with the temperature value. These informations are then directly used as input in the SOH estimation model, producing a SOH estimation. This estimation step is done in real time, and the calculus time is highly dependent to the variety of training datasets. Indeed, the estimation process require the quantification of DTW distance between new extracted patterns and all of the corresponding training patterns. However, this implementation is done in a few seconds and permits the application in real time. Figure 4 illustrates the obtained performances by this methodology with a battery under real EV use.

The global error between the estimations $\{\hat{h}_i | i = 1, ..., N\}$ and their respective targets $\{h_i | i = 1, ..., N\}$ is calculated with the relative error as follows :

$$\eta = \frac{1}{N} \sum_{i}^{N} \left| 1 - \frac{\hat{h}_i}{h_i} \right| \tag{22}$$

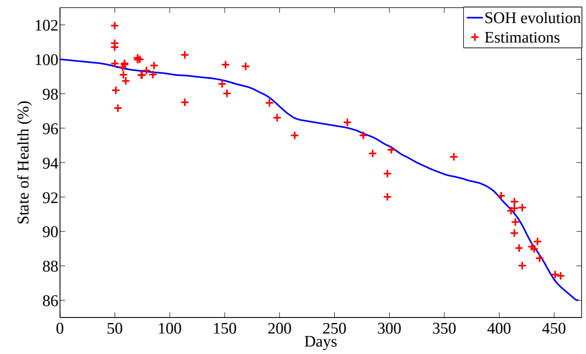



Figure 4. Estimation of the battery SOH in real-time, with a mean corresponding relative estimation error of $\eta = 0.81\%$

The SOH estimations demonstrate the good accuracy of the proposed methodology as the average relative error $\eta$ of the results illustrated in the Figure 4 is $0.81\%$. Note that the standard deviation of the obtained errors is $1.1\%$, inducing an interesting stability of the estimations. Moreover, the estimations trend fits with the SOH measurements, validating the reliability of this new innovative framework. Thus, Figure 4 shows that the use of machine learning process with battery patterns allows the estimation of the battery SOH. This result level is highly interesting as it performs to estimate the battery SOH with a good accuracy in real time during EV uses.

Based on these SOH estimations, the methodology detailed in Section 2.4 permits the prediction of the battery RUL at different times. Figure 5 presents the predicted battery RUL probability density functions at three different times, based on the SOH estimations illustrated in Figure 4.

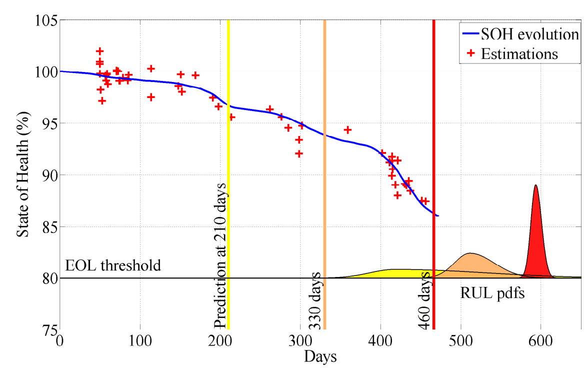



Figure 5. Prediction of the battery RUL probability density function (pdf) at three different times

The three RUL's probability density functions (pdf) presented in Figure 5 demonstrate the robustness of the proposed methodology. The chosen times for RUL estimation are made artificially to illustrate the evolution of the RUL pdfs over time.

Thus, the predicted RUL pdfs differ depending on the predicted time. The RUL pdf is indeed sensitive to the SOH estimation variations. For example we can see the RUL pdf predicted at 210 days considers a rapid battery capacity decrease, due to the last SOH estimations made before the prediction time.

It is also noticeable that RUL prediction improving in both accuracy and precision with the inclusion of more measurements before prediction. This is clearly visible in Figure 5, as the RUL pdf predicted at 460 days produces the best confidence level compared to the two others predictions. Thus, at 460 days the EOL criteria is predicted to be reached between 580 and 610 days, which represents a narrow range considering the total battery life.

## 4. CONCLUSION

This paper presents the implementation of a machine learning framework that allows the estimation of the battery State of Health (SOH) and predicting the Remaining Useful Life (RUL), and more specifically for Lithium-ion batteries. The proposed approach is based on the alteration of the battery signals behavior throughout its life to estimate the battery SOH, adapting the value of unknown model parameters during a preliminary training process. The estimated value of the battery capacity is then used to predict the battery RUL. Implementation results show the robust performance of the algorithm in real-time SOH estimation under uncontrolled conditions. The presented method performs well in a real life context, which is not the case of other existing approaches.

This study developed an innovative approach devoted to estimate the battery SOH during real EV uses, without specific requirements. Such a methodology is a particular advantage for a commercial aspect as it does not require to control the battery life conditions to make an estimation. The learned algorithm can indeed be used for estimating the SOH of all batteries with the same design. Meaning that once the estimation model is built, it can be used as an embedded estimation model in all EVs.

Furthermore, the average estimation error of less than $1\%$ obtained in the example presented in Figure 4 can be reduced using more training data coming from different EVs. In a fleet context, an estimation model can be trained from several EVs and then be embedded into all EVs using the same battery. This would allow accurate SOH estimation in all of these EVs.

Data driven approaches require large dataset to perform, however the results presented were obtained from a model built on a single EV. This study demonstrates a new baseline for SOH estimation only based on battery signals. It would be interesting to use this algorithm with a large set of data coming from an EV fleet. Thus, the next step is to test this methodology with several batteries to demonstrate the robustness and accuracy of the developed process with a large EVs fleet. In this case, the learned model by machine learning process would deliver even more accurate SOH estimations, as it would be based on more training datasets. To extend this methodology, it can also be considered in future studies to explore new kernels associations in order to input more information to the machine learning step.

The presented methodology can also be transposed to every battery use, to estimate its SOH during real utilizations. Indeed, this study does not consider any restricting use hypothesis. This methodology is, for example, adjustable in electric aircraft context.

## REFERENCES

Aach, J., & Church, G. M. (2001). Aligning gene expression time series with time warping algorithms. *Bioinformatics*, *17*(6), 495-508.

Bahlmann, C., Haasdonk, B., & Burkhardt, H. (2002). Online handwriting recognition with support vector machines - a kernel approach. In *Frontiers in handwriting recognition, 2002. proceedings. eighth international workshop on* (p. 49-54).

Bar-Joseph, Z., Gerber, G., Gifford, D. K., Jaakkola, T. S., & Simon, I. (2002). A new approach to analyzing gene expression time series data. In *Proceedings of the sixth annual international conference on computational biology* (pp. 39–48). New York, NY, USA: ACM.

Barré, A., Deguilhem, B., Grolleau, S., Gérard, M., Suard, F., & Riu, D. (2013). A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. *Journal of Power Sources*, *241*(0), 680 - 689.

Barré, A., Suard, F., Gérard, M., Montaru, M., & Riu, D. (2014). Statistical analysis for understanding and predicting battery degradations in real-life electric vehicle use. *Journal of Power Sources*, *245*(0), 846 - 856.

Lei, H., & Sun, B. (2007). A study on the dynamic time warping in kernel machines. In *Signal-image technologies and internet-based system, 2007. third international ieee conference on* (p. 839-845).

Peters, A., & Dütschke, E. (2014). How do consumers perceive electric vehicles? a comparison of german consumer groups. *Journal of Environmental Policy & Planning*, *0*(0), 1-19.

Petitjean, F., Kurtz, C., Passat, N., & Ganarski, P. (2012).

Spatio-temporal reasoning for the classification of satellite image time series. *Pattern Recognition Letters*, *33*(13), 1805 - 1815.

Saha, B., & Goebel, K. (2008). Uncertainty management for diagnostics and prognostics of batteries using bayesian techniques. *Aerospace Conference, IEEE*, 1-8.

Sakoe, H., & Chiba, S. (1971). A dynamic programming approach to continuous speech recognition. In *Proceedings of the seventh international congress on acoustics, budapest.*

Suard, F., & Mercier, D. (2009). Using kernel basis with relevance vector machine for feature selection. In *Artificial neural networks icann 2009* (Vol. 5769, p. 255-264). Springer Berlin Heidelberg.

Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.*, *1*, 211–244.

## BIOGRAPHIES

**Anthony Barré** is a Ph.D. student in applied statistics at the French Atomic Center (CEA), France. His research interests include complex data analysis, the study of signals behavior depending on performances level and machine learning applications.

**Frédéric Suard** obtained his Ph.D. degree in 2006 at the National Institute of Applied Sciences (INSA) in machine learning and computer science. Since 2007, he is a research engineer at the CEA, in charge of both industrial application and collaborative research projects mainly devoted to data analysis and modeling.

**Mathias Gérard** received his Ph.D. degree in 2010 from the CEA and Fuel Cell Laboratory in Belfort, France. Since 2013, he supervises the fuel cell and battery modeling activities at CEA. His research interests include electrochemical nano and micro scales, ageing phenomena, energy management strategies and control laws, including fast prototyping.

**Delphine Riu** received her Ph.D. degree in 2002 and her Habilitation in 2011 in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), France. She is currently a Professor at the Grenoble INP, in Grenoble Electrical Engineering lab. Her current research interests include modeling of power sources, energy management and design of hybrid electrical systems.