

# Comparison of Stochastic Capacity Estimation Tools Applied on Remaining Useful Life Prognosis of Lithium Ion Batteries

Mikel Arrinda<sup>1</sup>, Mikel Oyarbide<sup>2</sup>, Haritz Macicidor<sup>3</sup>, Eñaut Muxika<sup>4</sup>

<sup>1,2,3</sup> *CIDETEC, Donostia, Gipuzkoa, 20014, Spain*

*marrinda@cidetec.es*

*moyarbide@cidetec.es*

*hmacicior@cidetec.es*

<sup>4</sup> *Mondragon Unibertsitatea, Arrasate, Gipuzkoa, 20500, Spain*

*emuxika@mondragon.edu*

## ABSTRACT

Lithium ion batteries suffer a performance decrease in normal use, which leads to an end to the usability of those batteries under the defined application. Accurate estimation of the useful lifetime of the batteries is important in order to achieve high energy efficiency and cost reduced designs. However, the aging of a lithium ion battery has a non-linear behavior and the models available nowadays are far from completely describing it. Thus, estimation algorithms are applied in order to improve the accuracy of current models. In this sense, this paper evaluates different stochastic tools applied to typical capacity fade models in order to increase the accuracy of the useful life estimations at a specific aging state. The chosen capacity fade models are semi-empirical models based on power laws, exponentials and polynomials which represent in a simple way the capacity decrease of lithium ion cells under specific conditions. The degradation data used in this paper is obtained from LiFePO<sub>4</sub>-graphite (LFP) and LiNi<sub>0.8</sub>Co<sub>0.1</sub>Al<sub>0.1</sub>O<sub>2</sub>-graphite (NCA) aged cells. Particle Filter (PF) and Gaussian Process Regression (GPR) based stochastic algorithms are applied to improve especially the End of Life estimation. The comparison of the algorithms is performed based on a innovative comparison framework that discriminates external uncertainty effects. The benefits and limitations of the algorithms are quantified by the relevant metrics defined in the comparison framework. As a result, some guidelines on key aspects on the design and implementation of the algorithm are provided.

## 1. INTRODUCTION

Due to their outstanding properties, lithium ion batteries

(LIBs) are one of the most used energy storage devices in different applications such as portable electronic devices, electric vehicles and stationary back-up energy storage systems. However, the consequences of the failure of a battery can have different levels of severity on those applications ranging from reduced performance to non-fulfilment of the operational requirements to even catastrophic failures. An efficient method for battery monitoring would greatly improve the reliability of such systems (Goebel, Saha, Saxena, Celaya, & Christophersen, 2008), which will lead to the suitable sizing of the system and to a reduction on its cost (high energy efficiency and cost reduced design).

Nonetheless, the fact is that the monitoring or the prediction of the characteristics of the battery along its lifespan (prediction of the degradation and the remaining useful life (RUL)) done with certainty is not a trivial issue (Si, Zhang, & Hu, 2016). Firstly, it is almost impossible to observe the battery internal electrochemical process. Secondly, the aging of a LIB is a non-linear and time variable system (Rezvanizani, Liu, Chen, & Lee, 2014). Thirdly, environmental uncertainties affect the production and the performance of these LIBs (dynamic environments induce changes in the physics of failure (Si et al., 2016)). Based on this, it could be stated that a reasonable and appropriate degradation model applied on a LIB RUL prognosis problem has to take into account uncertainty of battery behaviour and uncertainty of the internal characteristics (Rezvanizani et al., 2014).

To take into account the effect of those uncertainties in the battery monitoring, and more precisely in RUL prognostics, stochastic tools are commonly applied to aging models (Gorjian, Ma, Mittinty, Yarlalagadda, & Sun, 2009). In these cases, the RUL prediction is based on stochastic degradation processes and performance degradation data modelling. Thanks to the stochastic degradation process

---

Mikel Arrinda et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

implementation, the distribution of the RUL is inferred, which allows the quantification of the uncertainty of the predicted results (Wu, Fu, & Guan, 2016). There are many available stochastic algorithms that can be applied to LIB RUL prognostic problems, and additional ways of implementing each one on the problem at hand. However, there is not a clear way of deciding which one should be the chosen one in each case.

Motivated by the lack of guidelines on the selection of the suitable prediction tool, this paper pursues a detailed evaluation of some configurations of popular stochastic algorithms in order to help researchers on that selection step. Among the available stochastic algorithms, the analysed ones in this paper are the GPR (with different covariance functions) (Richardson, Osborne, & Howey, 2017), (Zhou et al., 2018) and the PF (with different resampling methods) (H. Zhang, Miao, Zhang, & Liu, 2018), (Duong & Raghavan, 2018).

The characteristics of the tested cells and the applied aging models are explained in Section 2. The used stochastic tools (GPR and PF) are described in Section 3. The comparison approach is summarized in Section 4. The obtained results are shown in Section 5. The discussion and description of the results are presented in Section 6. Finally, Section 7 contains the conclusions and future work proposals.

## 2. LITHIUM ION BATTERY AGING MODEL

The proposed aging model is based on a semi-empirical capacity decrease model in terms of available dischargeable capacity. The goal of the proposed method is to describe the aging behaviour and estimate the End of Life (EOL) in terms of available dischargeable capacity as the result of a RUL prognosis problem. This is applied to degradation data obtained from 2 different LIB technologies (NCA and LFP).

### 2.1. NCA cell

The aging data from the NCA cell has been taken from NASA's data repository (Saha & Goebel, 2007). Among the different available data repositories, the battery "B0005" is the chosen one due to its length and because the same battery was used in recent studies (Duong & Raghavan, 2018) (H. Zhang et al., 2018).

The selected NASA dataset consists on a rechargeable 18650 Gen 2 Li-ion cell with a rated capacity of 2Ah. The experiment was conducted through three different operational profiles (charge, discharge and impedance) at room temperature. Charging was performed in a constant current at 1.5A until the battery voltage reached 4.2V and continued in constant voltage mode until the charge current dropped to 20mA. The discharge runs were stopped at 2.7V. The experiments were conducted until the capacity decreased to the specified EOL criteria of 1.4Ah (Tao et al., 2017).

(Saha, Goebel, & Christophersen, 2009) and (Goebel et al., 2008) used an exponential growth model shown in Eq. (1), where  $\theta$  represents the internal battery model parameter such as the available discharge capacity in each state of health,  $t$  represents the amount of cycles done and  $C$  and  $\lambda$  represents the relevant decay parameters.

$$\theta = C \exp(-\lambda t) \quad (1)$$

In contrast to the previous model, many other authors (Duong & Raghavan, 2018), (H. Zhang et al., 2018), (Chen & Pecht, 2012) proposed the capacity decrease model shown in Eq. (2) for the collected aging data in the NASA's data repository.  $a, b, c, d$  represent the inner parameters that needs to be fitted,  $t$  is the amount of cycles done and  $cap$  is the available discharge capacity in each cycle.

$$cap(t) = a \cdot \exp^{b \cdot t} + c \cdot \exp^{d \cdot t} \quad (2)$$

On the other hand, according to (Saha & Goebel, 2009), in order to effectively determine the End of Life (EOL) of a LIB, we need to understand how the different operational modes, namely charge, discharge and rest, influence the discharge capacity. The model shown in Eq. (3) is based partly on their proposal, where the self-recharge during test is represented as an exponential process as suggested by data (dependant on the rest period between cycles  $k$  and  $k+1$  ( $\Delta t_k$ ) and the cumulative capacity fade between rest periods ( $\Delta C_k$ )) and the capacity fade is described by the reduction of the previous discharge capacity ( $C_k$ ) by a constant ( $\beta_3$ ).

$$C_{k+1} = C_k + \beta_1 \exp(-\beta_2 / \Delta t_k) \Delta C_k - \beta_3 \quad (3)$$

### 2.2. LFP cell

The aging data from the LFP cell has been taken from CIDETEC's own database. This cell consists on a homemade cell with a rated capacity of 15Ah. The experiment was part of the test plan matrix shown in Table 1. Among them, the data set that holds more data has been chosen: the data from the test line n° 7.

Table 1. Test matrix design aimed to evaluate Cycling life.

Test	T° [°C]	SOC <sub>ini</sub> [%]	DoD [%]	C <sub>rate</sub> [-]
1	15	90	70	1
2	15	80	50	2
3	15	70	30	3
4	30	90	50	3
5	30	80	30	1
6	30	70	70	2
7	45	90	30	2
8	45	80	70	3
9	45	70	50	1

The test plan in test line n° 7 comprised a periodically cycling and measurement test: after cycling 300 cycles at the defined conditions, the measurement test was performed. The cycling test comprises a discharge at constant current of 1C (15A) at 45°C, starting at a 90% initial State of Charge ( $SOC_{ini}$ ) and ending once a 30% Depth of Discharge (DOD) is fulfilled. The charge is performed at constant current of 2C (30A) at 45°C until the 90%  $SOC_{ini}$  is reached. The measurement test comprised standard capacity, resistance and power measurements at 25°C.

The capacity fade model applied to the data of interest is based on a power law shown in Eq. (4), where  $Q$  represents the available dischargeable capacity obtained in the measurement tests,  $Ah$  represent the discharged capacity along the cycling tests and  $P$  is the power law that contains the effect of the tested stress factors on the cycling tests (temperature ( $T^\circ$ ), initial State of Charge ( $SOC_{ini}$ ), Depth of Discharge (DOD) and current rate ( $C_{rate}$ )).

$$Q = Ah^P \quad (4)$$

### 3. STOCHASTIC TOOLS

The stochastic tools manage parameter sets of chance variables (Doob, 1934). Thanks to the stochastic tools, the uncertainty can be taken into account, as well as the historical evaluation of the data. The Particle Filter and the Gaussian Process Regression are part of the available stochastic tools that can deal with those variables and that are present on today's studies of LIB RUL prognosis problems (Richardson et al., 2017), (Zhou et al., 2018), (H. Zhang et al., 2018), (Duong & Raghavan, 2018).

#### 3.1. Particle Filter

Particle Filter (PF) is a sequential Monte Carlo method, which estimates the state Probability Density Function (PDF) from a set of particles and their associated weights (J. Zhang & Lee, 2011). It is based on the idea of Monte Carlo method to solve the integral operation in the Bayes estimators. It estimates the state PDF from a set of "particles" and their associated weights (X. Zhang, Miao, & Liu, 2017). The use of weight adjusts the state PDF to its most likely form. Thanks to the use of state PDF, an appropriate management of inherent estimation uncertainty is allowed (J. Zhang & Lee, 2011). This provides non-linear projection in forecasting (Heng, Zhang, Tan, & Mathew, 2009).

The particles are inferred recursively by two alternate phases. The first phase is the prediction where the value of each particle for the next step is estimated by previous step information. No measurement or observation is involved in this step. The second phase is the update where the value of each particle estimated in the prediction phase is compared

with measurements and updated accordingly (J. Zhang & Lee, 2011).

In LIB RUL prognostics problems, the capacity estimate ( $\hat{Q}_{t+1}$ ) can be calculated by adding the product between the normalized weight of each particle  $j$  ( $w_{t+1}^j$ ) and the capacity estimated value obtained on each particle ( $Q_{t+1}^j$ ) described in Eq. (5) (Chen & Pecht, 2012).

$$\hat{Q}_{t+1} = \sum_{j=1}^{N_j} w_{t+1}^j Q_{t+1}^j \quad (5)$$

The same proposal can be applied to estimate the states ( $\hat{\theta}_{t+1}$ ) using the state of each particle ( $\theta_{t+1}^j$ ) (see Eq. (6)) (Arulampalam, Maskell, Gordon, & Clapp, 2002). After this, the capacity can be calculated using the equations of the capacity decay model shown in the previous Section.

$$\hat{\theta}_{t+1} = \sum_{j=1}^{N_j} w_{t+1}^j \theta_{t+1}^j \quad (6)$$

In both cases, the weight is updated with the posterior PDF. Since the posterior PDF ( $P(Q_{t+1}|y_{0:t+1})$  or  $P(\theta_{t+1}|y_{0:t+1})$ ) is usually unknown, importance sampling principle is used to sample  $Q_{t+1}^j$  or  $\theta_{t+1}^j$  from an importance density ( $q(Q_{t+1}|Q_t, y_{0:t+1})$  or  $q(\theta_{t+1}|\theta_t, y_{0:t+1})$ ), and the corresponding weights can be updated by the Eq. (7).

$$w_{t+1} \propto w_t^j \frac{P(y_{t+1}|Q_{t+1})P(Q_{t+1}|y_t)}{q(Q_{t+1}|Q_t, y_{0:t+1})} \quad (7)$$

The prior PDF of the capacity can be described by the Eq. (8), where the capacity posterior PDF of the capacity can be calculated by Bayes' rule like in the Eq. (9).

$$P(Q_{t+1}|y_{0:t}) = \int P(Q_{t+1}|Q_t)P(Q_t|y_{0:t})dQ_t \quad (8)$$

$$\begin{aligned} P(Q_{t+1}|y_{0:t+1}) &= \frac{P(y_{0:t+1}|Q_{t+1})P(Q_{t+1})}{P(y_{0:t+1})} \\ &= \frac{P(y_{t+1}|Q_{t+1})P(Q_{t+1}|y_{0:t})}{P(y_{t+1}|y_{0:t})} \end{aligned} \quad (9)$$

If choosing the marginal likelihood of the posterior PDF, the update of the weights can be reduced to Eq. (10), and thus, reduced to Eq. (11).

$$w_{t+1} \propto w_t^j P(y_{t+1}|Q_{t+1}) \quad (10)$$

$$w_{t+1} \propto w_t^j \delta(y_{t+1} - y_{t+1}^j) \quad (11)$$

Similarly, if choosing the marginal likelihood, the posterior PDF can be simplified to Eq. (12).

$$P(Q_{t+1}|y_{0:t+1}) \approx P(y_{t+1}|Q_{t+1}) \approx \sum_{j=1}^{N_j} w_t^j \delta(Q_{t+1} - Q_{t+1}^j) \quad (12)$$

Between the two proposals available on the literature, the PF algorithms developed for this study estimate the inner states of the capacity fade models (the fitting parameters on the models). Besides, for calculus, instead of using the delta of Dirac ( $\delta$ ), a Gaussian distribution function is used. In this way, the neighbor values are praised instead of praising only the equal values and despise the rest. On the same line, the fact is that an initialization set artificially could improve the tracking ability of the PF based prognostic method (Wang, Yang, Zhao, & Tsui, 2017). This paper proposes an initialization that assumes the first state of every particle to be randomly normally distributed and the first weights to be uniformly distributed.

Among the disadvantages of this stochastic tool, there are two main problems: Particle degradation and sample impoverishment (X. Zhang et al., 2017). In order to reduce them, system importance resampling of the particles is commonly carried out on each iteration that the pre-set resampling threshold is not reached. This helps in maintaining the track of the state vector even under the presence of disruptive effects like un-modelled operational conditions (Goebel et al., 2008). In the literature, the most common approaches are the basic Systematic Resampling (SR) (Arulampalam et al., 2002), the Multinomial Resampling (MR) and the Residual Resampling (RR) (Douc, Cappé, & Moulines, 2005). The algorithm of the three resampling methods are available in (Li, Bolic, & Djuric, 2015).

### 3.2. Gaussian Process and Gaussian Process Regression

The Gaussian Process (GP) is based on the statistical learning theory and adapts well to high dimensions, small samples, nonlinearities and other complex problems with a strong generalization ability (Wu et al., 2016).

In a GP, observations occur in a continuous domain (time or space) and every point is associated with a normally distributed random variable. This supposes that every finite collection of those random variables has a multivariate normal distribution and that every finite linear combination of them is normally distributed. Supported by those assumptions, a GP defines a probability distribution over functions (see Eq. (13)) which are composed by a mean function (Eq. (14)) and a covariance function (Eq. (15)). In this way, the degradation trends are learnt from battery data sets with the combination of GP functions.

$$f \sim GP(m(x), k(x, x')) \quad (13)$$

$$m(x) = E(f(x)) \quad (14)$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x')))] \quad (15)$$

Typically a GP uses a mean function equal to zero with the aim of describing all the system by the covariance function. This configuration is typically adopted because the covariance function is flexible enough to model the true mean arbitrary well (Rasmussen, 2006). However, in case of having prior knowledge of the system such as the capacity decay model in LIB RUL estimation problems, it is interesting to express this prior information as the most probable result of the systems in form of the mean function. This configuration of GP is able to describe the uncertainties of the prior knowledge by the covariance function. The chosen configuration for this paper is the last one.

In Gaussian Process Regression (GPR), the distribution (over functions) obtained by the GP is used as a prior for Bayesian inference (Eq. (16)). The calculated prior does not depend on the training data, but specifies some properties of the functions (the objective is to learn properties of the prior in the light of the training data) (Rasmussen, 2006). The calculation of the posterior will provide the predictions for unseen test cases. Then, the joint distribution of the desired test set is evaluated where the training set covariance ( $K$ ), training-test set covariance ( $K_*$ ) and the test set covariance ( $K_{**}$ ) are calculated.

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right) \quad (16)$$

Since the values for the training set  $f$  are known, the conditional distribution of  $f_*$  given  $f$  can be calculated by Eq. (17) (this is the posterior distribution for a specific set of unseen test cases).

$$f_* | f \sim \mathcal{N}(\mu_* + K_*^T K^{-1}(f - \mu), K_{**} - K_*^T K^{-1} K_*) \quad (17)$$

In the same way, the mean and the variance of the posterior can be deduced from here (Eq. (18) and Eq. (19) respectively).

$$m_p(x) = m(x) + K_*^T K^{-1}(f - \mu) \quad (18)$$

$$k_p(x, x') = K_{**} - K_*^T K^{-1} K_* \quad (19)$$

The estimation attained with this Bayesian inference is noiseless, however, it is something common to have noise in the observations of regression applications. In the GP models, such noise is easily taken into account. The easiest way of adding the noise effect in the observation is to assume that the noise is Gaussian and independent. In this scenario, the noise variance is added to the covariance values of each test point respect to the same test point ( $k_p(x, x')$ ;  $x = x'$ ) (Eq. (20)).

$$K_y = K + \sigma_y I \quad (20)$$

To sum up, the GPR can model the behaviour of any system through the combination of the appropriate GP and prior knowledge. However, an appropriate GP means selecting an appropriate covariance function, which is an important and difficult problem (Long, Xian, Jiang, & Liu, 2013). Studies of LIB RUL prognosis problems have applied covariance functions such as the Squared Exponential (SE) covariance function (Eq. (21)), the Matérn (Ma) covariance function (Eq. (22)) (Richardson et al., 2017) and the neural network (NN) covariance function (Eq. (23)) (Zhou et al., 2018).

$$k_{SE}(x_p, x_q) = \sqrt{\pi} l \sigma_p^2 \exp\left(-\frac{(x-x')^2}{2(\sqrt{2}l)^2}\right) \quad (21)$$

$$k_{Ma}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}r}{l}\right) \quad (22)$$

$$k_{NN}(x, x') = \frac{2}{\pi} \sin^{-1}\left(\frac{2\tilde{x}^T \Sigma \tilde{x}'}{\sqrt{(1 + 2\tilde{x}^T \Sigma \tilde{x})(1 + 2\tilde{x}'^T \Sigma \tilde{x}')}}\right) \quad (23)$$

#### 4. COMPARISON FRAMEWORK

This paper aims at contributing on the selection of the appropriate stochastic tool applied on LIB RUL prognosis problems, in concrete those used on capacity fade models. However, in order to help on the stochastic tool selection process, a comparison of both tools needs to be done, so firstly, a proper comparison approach is required. The proposed comparison framework is based on several trials designed to minimize the induced sources of uncertainty of the applied inputs on the stochastic tools. Once run the trials, the merits of each tool are quantified by some defined key metrics.

##### 4.1. Design of the trials

The design of the trials needs to keep in mind that many of the sources of uncertainty on the RUL estimation are “inputs” to the prognostic algorithm. These uncertainties on the “inputs” can penalize the algorithm if the information regarding these “inputs” is incorrect; it would not be reasonable to penalize or accept an algorithm according to the fitness of the prediction respect to the ground truth data in case the algorithm did not have access to an accurate degradation model and/or an accurate estimate of the future conditions of the component/system (Sankararaman, Saxena, & Goebel, 2014). That is why it is necessary to develop a rigorous comparison approach to separate:

- The evaluation of correctness of information regarding these “inputs”.
- The evaluation of the prognostic algorithm itself.

This paper proposal aims to somehow control the uncertainty of the “inputs” when designing the trials, which will allow evaluating the prognosis algorithm itself. The idea is, firstly, to apply different prior knowledge describing the behaviour of the system (the different capacity decay models of the NCA cell proposed in the literature) and secondly, to apply data from another system that shares similarities but which is not the same (the data from the LFP cell). In this way, the correctness of the “inputs” for each stochastic tool can be discriminate, making possible the evaluation of the prognostic algorithms themselves.

Similarly, the prognostic algorithms called PF and GPR have many settings that change the obtained results. Our goal is to compare the PF and the GPR, but since their results depend on the used configuration (such as the resampling method on PF and the covariance function on GPR), different recent settings of these prognostic algorithms are evaluated. The PF is set with 3 common resampling methods: SR, MR and RR; and the GPR is set with 3 different covariance functions: SE, Ma and NN.

The whole test matrix is shown in Table 2.

Table 2. Trials of the stochastic tools

Test n°	Stochastic tool	Cell	Aging model	Differential characteristic
1	Particle Filter	NCA	Eq. (1)	SR
2				MR
3				RR
4			Eq. (2)	SR
5				MR
6				RR
7		Eq. (3)	SR	
8			MR	
9			SR	
10		LFP	Eq. (4)	SR
11				MR
12				RR
13	Gaussian Process Regression	NCA	Eq. (1)	SE covariance
14				Ma covariance
15				NN covariance
16			Eq. (2)	SE covariance
17				Ma covariance
18				NN covariance
19		Eq. (3)	SE covariance	
20			Ma covariance	
21			NN covariance	
22		LFP	Eq. (4)	SE covariance
23				Ma covariance
24				NN covariance

In the process of running all the trials defined above (Table 2), some parameters have been maintained constant to be able to compare both prognostics tools and their different configurations in equal terms. These parameters are resumed in Table 3.

Table 3. Affecting parameters.

Parameters	Value	Description	Used in
N	1/2	Training data set.	PF + GPR
L	1/4	Validation data set.	PF + GPR
$EOL_{NASA}$	1.3182	Capacity value at 7/8.	PF + GPR
$EOL_{LFP}$	80.456	Capacity value at 7/8.	PF + GPR
$p_{kop}$	1000	Particle quantity.	PF
$Neff$	50%	Resampling threshold.	PF

## 4.2. Comparison Metrics

Once designed and run the trials, the obtained results need to be quantified in order to be compared. The proposed comparison focuses on 4 aspects described by the metrics resumed in Table 4:

- The improvement on the fitting of the model.
- The error on the desired state estimation.
- The uncertainty level of the estimations.
- The computing load of the tool.

Firstly, the ability of the tool to adapt to the data of the tested system is checked. This is quantified by the Root Mean Squared Error (RMSE) (Eq. (24)) on the whole data set, differencing the error obtained on the training data set from the error obtained on the prediction data set.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (24)$$

Secondly, the EOL estimation error is evaluated. The Relative Accuracy (RA) of the predicted EOL is checked (Saxena, Celaya, Saha, Saha, & Goebel, 2010). The metric is described in Eq. (25), where the  $EOL_m$  is the measured EOL value and  $EOL_e$  is the estimated EOL value.

$$RA = 1 - \frac{|EOL_m - EOL_e|}{EOL_m} \quad (25)$$

Thirdly, the uncertainty level of the estimation is evaluated through measurements on the PDF. The probability of estimating the real EOL by the tool is used to know if the uncertainty is under-estimated and the PDF width is used to check in contrast, if the uncertainty is over-estimated. The PDF calculation on each algorithm is evaluated in a different way. The PDF in the algorithms based on GPR is defined by the variance achieved in the prediction step and the joint Gaussian distribution property of GP. The PDF in the algorithms based on PF is defined by the density distribution of the RUL estimations of each particle obtained at the last learning step.

Fourthly, the computing load of the tool in the computer is measured by the computing time per training data point. For this, the mean time of the obtained time in 5 runs is measured.

Table 4. Description of the comparison metrics.

Metric	Description
Training RMSE	Root Mean Squared Error (RMSE) on the learning data set (ordinate).
Prediction RMSE	Root Mean Squared Error (RMSE) on the prediction data set (ordinate).
RA	Relative Accuracy (RA) of the predicted EOL respect to the measured EOL (abscissa).
EOL P value	The probability of estimating the measured EOL value (from a normalized PDF).
PDF width	The width of the PDF in the ordinate axis with a 68% confidence range.
Computing time / data	The computing time of generating the prediction divided by the length of the prediction data set.

## 5. RESULTS

The parameters used in the evaluated algorithms are achieved by grid search under restrained conditions defined in Table 3. The created grids are filled with logarithmic stepped values. Once created the grids, the optimum parameters for each algorithm's configuration have been chosen by a cross validation on the defined validation data set. Then, the trials described in Table 2 have been run.

The achieved prognosis results in all the trials are shown in Figure 2. The legends of each object on Figure 2 have been removed and shown in Figure 1 because they do not allow the proper visualization of the results and they are repeated in every chart.

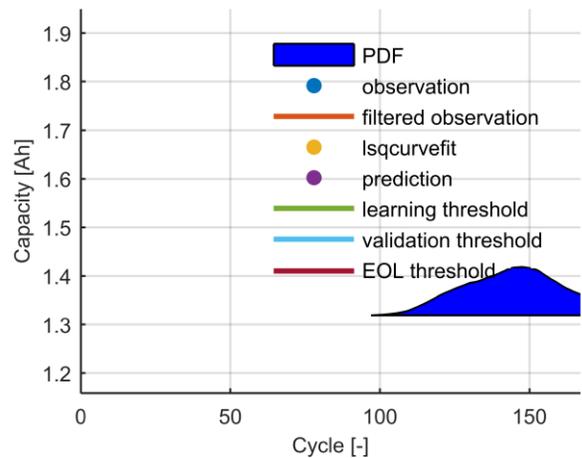
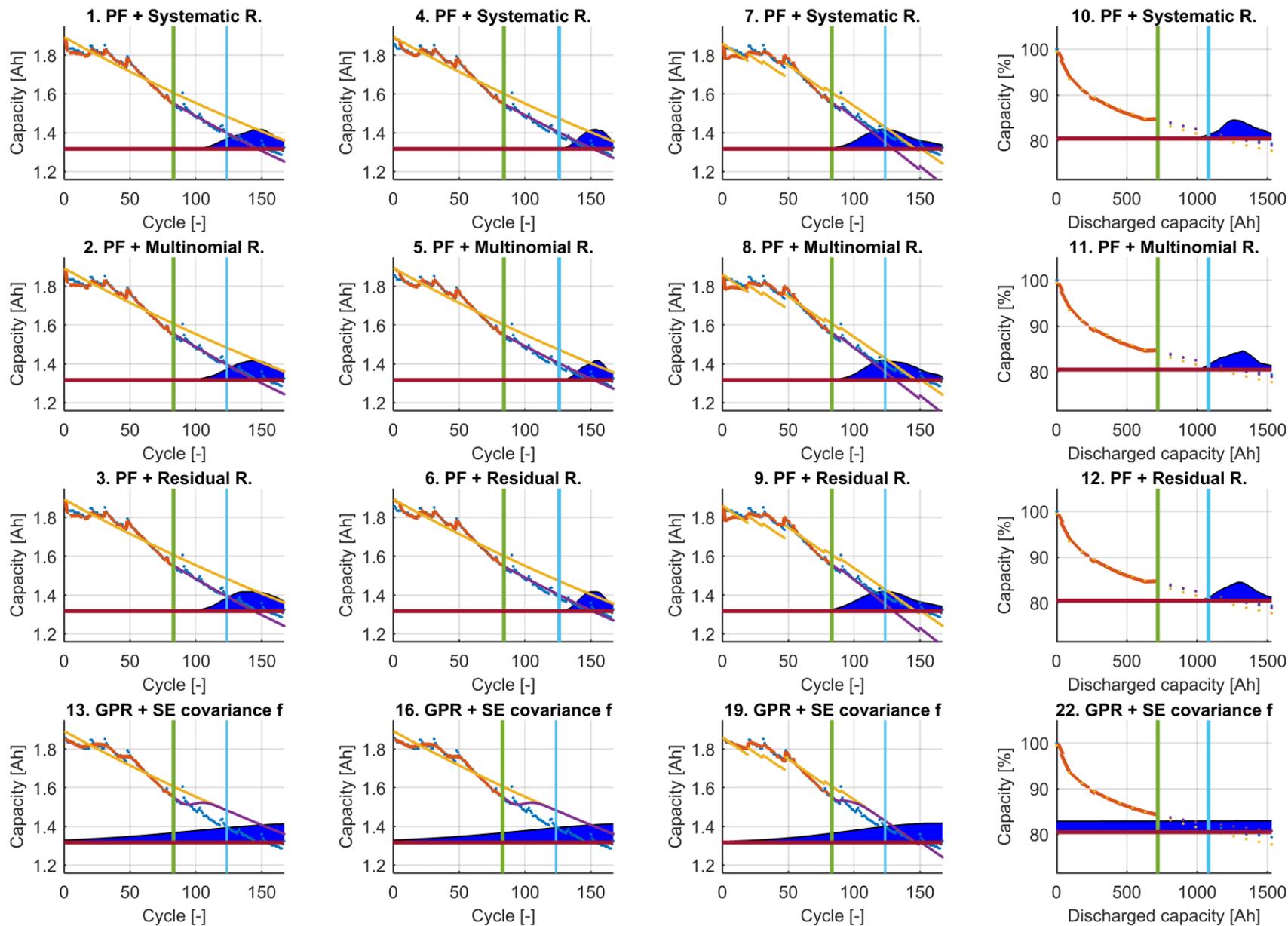


Figure 1. Legend of the results obtained in each trial.

Finally, the metrics used for comparison described in Table 4 have been calculated and written in Table 5.



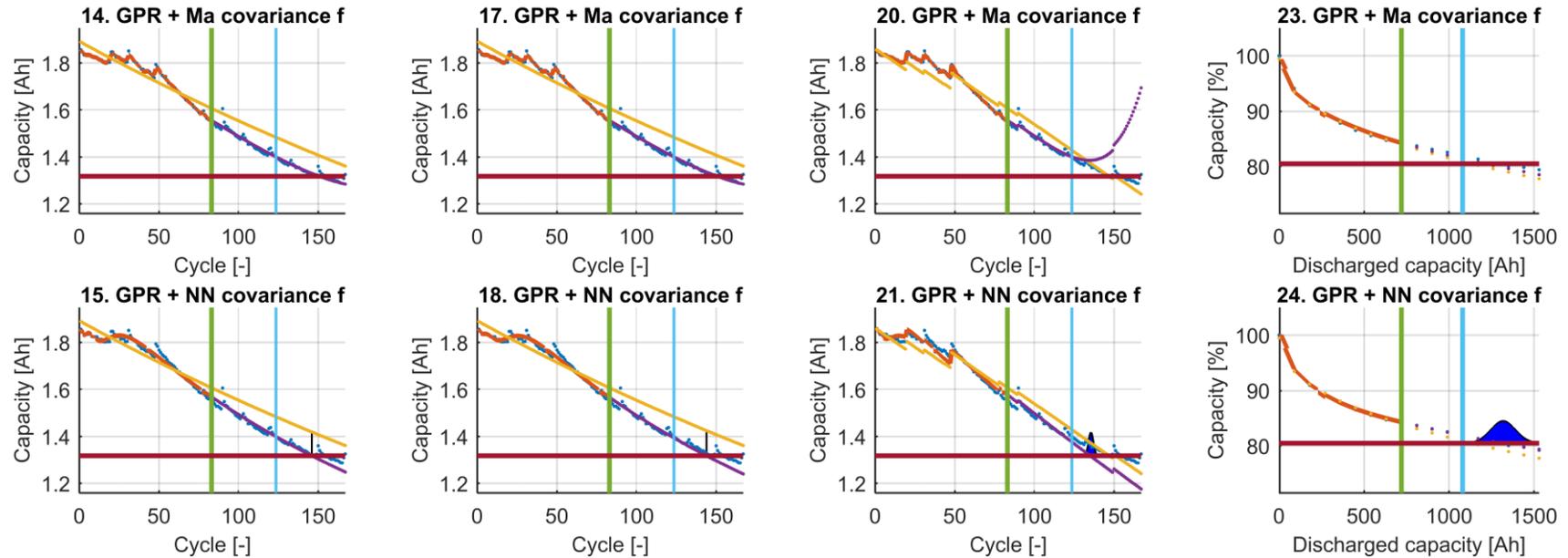


Figure 2. Results of the trials defined in Table 2. The dark blue is the PDF; the blue points are the measurement of the available dischargeable capacity (observation); the red line is the fitting capability of the applied prognostic tool (filtered observation); the yellow points are the fitting of the learning data set with MATLAB's lsqcurvefit function (lsqcurvefit); the purple points are the future estimation calculated by the prognostic tool (prediction); and the rest 3 lines represent different threshold (green is the end of the training data set, blue is the end of the validation data set end, red is the EOL value).

Table 5. Metrics used to compare the prognosis tools.

Comparison Metrics	Eq. (1)						Eq. (2)						Eq. (3)						Eq. (4)					
	PF			GPR			PF			GPR			PF			GPR			PF			GPR		
	SR	MR	RR	SE	Ma	NN	SR	MR	RR	SE	Ma	NN	SR	MR	RR	SE	Ma	NN	SR	MR	RR	SE	Ma	NN
Test n°	1	2	3	13	14	15	4	5	6	16	17	18	7	8	9	19	20	21	10	11	12	22	23	24
Training RMSE	0.00 304	0.00 278	0.00 319	0.00 892	0.00 430	0.01 681	0.00 209	0.00 193	0.00 197	0.00 892	0.00 430	0.01 610	0.01 282	0.01 340	0.01 249	0.00 410	0.00 140	0.01 718	0.07 764	0.06 691	0.12 235	0.37 549	0.53 603	0.34 715
Prediction RMSE	0.02 048	0.02 323	0.02 511	0.07 325	0.01 442	0.02 207	0.01 723	0.01 739	0.01 718	0.07 325	0.01 442	0.02 545	0.07 725	0.06 998	0.07 561	0.03 148	0.12 230	0.05 376	0.25 791	0.25 275	0.27 888	0.95 824	0.62 522	0.19 901
RA [x10-3]	6.85	0	6.84	260	34.2	0	41.1	47.9	41.1	260	34.2	13.7	109	95.9	109	6.84	NAN	75.3	19.0	20.6	12.7	257	36.5	38.1
EOL P value	0.00 308	0.00 335	0.00 268	0.00 368	0	124 5.23	0.00 188	0.00 231	0.00 252	0.00 368	0	0	0.00 234	0.00 284	0.00 228	0.00 554	0	6.08 E-12	0.00 314	0.00 264	0.00 259	6.48 E-06	0	0.00 395
PDF width	8.28 8	13.7 78	15.6 22	189	Inf	0.00 064	13.6 2	14.9 76	12.2 83	189	Inf	0.00 072	6.70 8	5.35 9	16.6 32	137	Inf	3	120. 048	77.8 75	140. 279	751 50	Inf	147
Computing time / data	0.03 630	0.03 789	0.03 779	0.03 022	0.00 486	0.04 893	0.03 864	0.03 817	0.03 911	0.01 497	0.00 449	0.06 212	0.10 956	0.10 946	0.10 825	0.01 235	0.00 467	0.08 027	0.03 559	0.03 645	0.03 993	0	0.00 086	0.00 260

## 6. DISCUSSION

First of all, note that the achieved results aim at evaluating the prognosis tools. The fact is that the learning step is larger in the prognosis tools than in the aging models due to the validation step needed in the applied optimization method (grid search). Therefore, any comparison between aging models and prognosis tools in this case would be inappropriate. However, thanks to the variety of aging models, we could discriminate the effect of the aging models on the obtained results and achieve our goal of evaluating the tools themselves and their different configurations.

A graphical display of the results obtained in each trial is shown in Figure 2 and the relevant metrics described in Section 4 are shown in Table 5. Thanks to the relevant metrics, the different configurations of the algorithms can be evaluated quantitatively while supporting the discussion by the qualitative evaluation of the graphs. The evaluation of the algorithms is divided in 4 parts:

1) The fitting error of the estimations respect to the ground data is evaluated by the first 2 metrics (“Training RMSE” and “Prediction RMSE”). According to the obtained values, the trial where the training data set is fitted the best is the trial n° 20 and the worst is the trial n° 23 (both run with the GPR + Ma covariance f). On the other hand, the best results in the fitting of the prediction data set (partially used in the cross validation) are obtained in the trial n° 14 and n° 17 (both run with the GPR + Ma covariance f) and the worst in the trial n° 22 (GPR + SE covariance f). This results show that the GPR is both, the best and the worst option when fitting the available data set on prognosis problems parametrized by grid search and cross validation.

Going further on this evaluation, it can be seen that the PF shows better results than the average except for the results obtained with the Eq. (3) (in both, in training and in prediction data sets), where the results obtained with the GPR are the ones under the average value. Checking the metrics on the graphs, it can be seen that the algorithms based on PF have inaccurate estimations on the first data points of the training data set. These first fitting errors increase the error of the fitting metric on the training data set, which explains the fact of not achieving the best results. However, it can also be seen that once these algorithms achieve a proper estimation, the next estimations fit accurately the rest training data set. This explains why the obtained fitting values are under the average. This applies to all four cases except for one case, where the obtained values surpass the average. The results obtained with Eq. (3) show that the fitting of the training data set as well as the fitting of the prediction data set of the algorithms based on PF are not within the best. The fact is that the proposed model in Eq. (3) describes the capacity fade by a constant, but it is not a

constant. The system shows a decrease on the capacity fade rate. This supposes that the model by itself is not able to fit well the data. In the case of the algorithms based on PF, the predictions are only based on the model parametrized with the last learned inner states. This means that the achieved fitting errors are as low as the best achievable by the model itself. This cannot be improved. On the other hand, the algorithms based on GPR are able to add to the prior model un-modelled behaviors of the system (learned on the training step). This leads to a more complete model that can get and gets lower fitting error values on the validation and prediction data set. However, it can also be seen that the trial n° 20 gets the higher fitting error on the prediction data set, even though being an algorithm based on GPR. This algorithm does not perform well on the prediction data set even though holding a validation fitting error under the average (used on the parametrization and validation of the algorithm). This type of behavior appears when an overfitting is done. The tool is over-fitted on the validation step and that leads to inaccurate future predictions, increasing considerably the prediction fitting error.

2) The error on the estimation of the defined EOL (the capacity value at 7/8 of the data set) is evaluated by the Relative Accuracy (RA). According to the obtained values, the trials with the best RA on the defined EOL value are the n° 2 (PF + MR) and n° 15 (GPR + NN covariance f). In this case, the exact prediction has been achieved with both algorithms. The worst relative accuracy is achieved on the trial n° 20, where the result is never predicted (“NAN”, not a number). As explained before, the algorithm applied on the trial n° 20 is over-fitted, leading to wrong prediction values. Additionally, it can be seen that in average, the worst values are achieved on trials n° 13, n° 16 and n° 22 (GPR + SE covariance f). Checking this on the graphs, it can be seen that the prediction done by the GPR with the SE covariance function starts losing the learned effect and ends making predictions based uniquely on the model itself. This means that the obtained RA values in those cases are the same as the RA values obtained by the models themselves.

Extending and deepening the analysis, it can be seen that the algorithms based on PF show RA values with the same magnitude and below the average. This means that in general, the algorithms based on PF are above the GPR algorithms in terms of RA values. However, it can be seen that the GPR with NN covariance function achieves the best RA values with the NCA battery data, but the worst RA value with the LFP battery data. Two of the possible reasons behind this are: (1) the available amount of data (not enough to learn the un-modelled behaviour of the system) and (2) the adequacy of the hyper-parameters or the correctness of the data. In this case, the proposed comparison approach has not been able to discriminate the uncertainty of the “inputs” and it cannot be known why the RA values are the best in 3 of 4 cases and the worst in the other case, but it gives hints on the potential of this algorithm.

3) The uncertainty level of the estimations is evaluated by the analysis of the PDF width and the probability of estimating the real EOL value (“EOL P value”). These two metrics along with the RA could be considered the most important metrics on prognostic tool evaluation. At a first glance, it could be seen that the configurations of GPR with Ma and SE covariance functions are not appropriate for uncertainty evaluation. The PDF obtained with these algorithms have width values higher than the real EOL value itself. The uncertainty is overestimated in those cases. The configuration of the GPR with NN covariance function, on the other hand, has such low PDF width values in three of the four tested cases that the probability on estimating the real EOL value gets 0 unless an exact prediction is done such as in the trial n° 15. The uncertainty is underestimated in those cases. The results obtained by the algorithms based on PF show differences when the “inputs” change but not big ones: the probability of estimating the real EOL remains similar in all the trials and looking at the graphs, it looks like the PDF widths are not under-estimating the uncertainty. In this case, it is not possible to confirm if the values estimate properly the uncertainty since the correctness of these metrics depends on the specifications of the application, which are not available in this study. However, it is clear that the tested algorithms based on PF are better than the tested algorithms based on GPR in terms of uncertainty evaluation.

4) The computing load of the tool on the PC is evaluated by the computing time per training data point. In this case, there is practically no difference in the algorithms based on the PF in terms of data points, which means that there is a common linear relation of the training data points and the computing time on each configuration of the PF. Nevertheless, the algorithms based on GPR show big differences depending on the amount of training data points (between NCA data and LFP data), which means that the relation of training data points and the computing time could be exponential, leading to the curse of dimensionality. This is a key aspect in on-line applications. Among the configurations of the algorithms based on GPR, the one with NN covariance function requires more time than the others, so this would be the worst in terms on computing load.

Analysing the results as a whole, it could be seen that the PF works similar in every case and with all the tested configurations (resampling methods), showing the huge possibilities that give this stochastic filter in prognostic problems. In all cases, the EOL measurement is in the generated probability distribution and the relative accuracy does not overcome the 10%, which in one case goes down to a 0%. On the other hand, the tested GPR configurations cases have shown the best results in most of the evaluated metrics, but they also have shown the worst values in the critical ones (RA, PDF width and EOL P value), showing the potential of the algorithms based on GPR but also the deficiencies on the tested configurations.

## 7. CONCLUSION

In this paper, two prognostic tools (GPR and PF) with several different configurations are applied to different case data (aging data of NCA and LFP cells) and different prior knowledge (capacity fade models) in a LIB RUL prognosis problem. The chosen prognostic tools have been then compared quantitatively according to some interesting metrics (see Table 4) supported by a qualitative comparison of the graphs shown in Figure 2.

Taking into account that the key to get useful prognostics is not only an accurate remaining life estimation, but also an assessment of the confidence of the uncertainty estimation (Goebel et al., 2008), the best prognostic tool tested in this study would be any of the algorithms based on PF. The calculated PDFs have a relatively reduced width in all the trials and the real EOL value can be found inside the probability distribution with a confidence higher than 68%. On the other hand, the algorithms based on GPR could only underestimate or overestimate the uncertainty of the EOL prediction, which is important not to do so (Sankararaman et al., 2014). In case of the GPR with Se covariance function, the PDF width is too big to consider the uncertainty estimation as appropriate (PDF width values higher than the the RUL value itself). Besides, it predicts EOL values equal to the ones obtained with the capacity fade model alone, which suggests that this configuration would not be interesting to improve the prior knowledge in prediction applications. In case of the GPR using the Matérn covariance function, the PDF width in all the trials is too big (infinite) even though if in some cases, the EOL estimation values (RA values) are among the best ones. This means that the estimated EOL value cannot be believed. An example of this asseveration can be found on the results obtained in the trial n° 20 where the predictions never reach the EOL threshold (NAN). In case of the GPR with NN covariance function, the results are far away from the ones obtained with the algorithms based on PF like the other GPR configuration, but it is worth noting that this is the algorithm that shows more potential among all the tested ones. Even though this algorithm requires improvement on the PDF estimation (the estimated PDF width is too narrow and the real EOL values stays out of the estimated probability distribution on 2 of the 4 tested cases, having a probability lower than 32% of estimating the real EOL), it achieves the best RA results on 3 of the 4 tested cases. The improvement of this algorithm could be achieved by (1) an improvement on the covariance function by adding or multiplying a covariance function to the NN covariance function, (2) an improvement on the aging model, (3) a reduction of the given uncertainty by the data or (4) an improvement on the parametrization algorithm.

For future works, a more robust comparison framework will be developed which will try to discriminate completely the effect of the uncertainties on the “inputs” of the stochastic

tools while also improving the algorithm. In the same way, the “Prognosis Horizon” and the “ $\alpha$ - $\lambda$  performance” metrics will be added on a 5<sup>th</sup> evaluation of the results. Besides, a need of synthesizing the results is detected, which will be attended. Finally, improvements will be applied to lead to a better understanding of the potentials of the algorithms and to lead to an easier selection of the appropriate algorithm in LIB RUL prognosis problems.

#### NOMENCLATURE

<i>DOD</i>	Depth of Discharge
<i>EOL</i>	End of Life
<i>GP</i>	Gaussian Process
<i>GPR</i>	Gaussian Process Regression
<i>LFP</i>	LiFePO <sub>4</sub> -graphite cell technology
<i>LIB</i>	Lithium-ion battery
<i>Ma</i>	Matérn covariance function
<i>MR</i>	Multinomial Resampling
<i>NCA</i>	LiNi <sub>0.8</sub> Co <sub>0.15</sub> Al <sub>0.05</sub> O <sub>2</sub> -graphite cell technology
<i>NN</i>	Neural Network
<i>PDF</i>	Probability Density Function
<i>RMSE</i>	Root Mean Squared Error
<i>PF</i>	Particle Filter
<i>RR</i>	Residual Resampling
<i>RUL</i>	Remaining Useful Life
<i>SE</i>	Squared Exponential covariance function
<i>SOC</i>	State of Charge
<i>SR</i>	Systematic Resampling
<i>T°</i>	Temperature

#### REFERENCES

- Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188. <https://doi.org/10.1109/78.978374>
- Chen, C., & Pecht, M. (2012). Prognostics of lithium-ion batteries using model-based and data-driven methods. *Proceedings of IEEE 2012 Prognostics and System Health Management Conference, PHM-2012*. <https://doi.org/10.1109/PHM.2012.6228850>
- Doob, J. L. (1934). STOCHASTIC PROCESSES AND STATISTICS. In *Proceedings of the National Academy of Sciences of the United States of America* (pp. 376–379).
- Douc, R., Cappé, O., & Moulines, E. (2005). Comparison of Resampling Schemes for Particle Filtering. <https://doi.org/10.1109/ISPA.2005.195385>
- Duong, P. L. T., & Raghavan, N. (2018). Heuristic Kalman optimized particle filter for remaining useful life prediction of lithium-ion battery. *Microelectronics Reliability*, *81*(December 2017), 232–243. <https://doi.org/10.1016/j.microrel.2017.12.028>
- Goebel, K., Saha, B., Saxena, A., Celaya, J. R., & Christophersen, J. P. (2008). Prognostics in Battery Health Management. *IEEE Instrumentation & Measurement Magazine*, 33–40. <https://doi.org/10.1002/9781119994053>
- Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., & Sun, Y. (2009). A review on degradation models in reliability analysis. In *Engineering Asset Management* (pp. 28–30).
- Heng, A., Zhang, S., Tan, A. C. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, *23*(3), 724–739. <https://doi.org/10.1016/j.ymssp.2008.06.009>
- Li, T., Bolic, M., & Djuric, P. M. (2015). Resampling Methods for Particle Filtering. *IEEE SIGNAL PROCESSING MAGAZINE*, (May), 70–86.
- Long, B., Xian, W., Jiang, L., & Liu, Z. (2013). An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries. *Microelectronics Reliability*, *53*(6), 821–831. <https://doi.org/10.1016/j.microrel.2013.01.006>
- Rasmussen, C. E. (2006). Gaussian processes for machine learning. *International Journal of Neural Systems*, *14*(2), 69–106. <https://doi.org/10.1142/S0129065704001899>
- Rezvanizani, S. M., Liu, Z., Chen, Y., & Lee, J. (2014). Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility. *Journal of Power Sources*, *256*, 110–124. <https://doi.org/10.1016/j.jpowsour.2014.01.085>
- Richardson, R. R., Osborne, M. A., & Howey, D. A. (2017). Gaussian process regression for forecasting battery state of health. Retrieved from <http://arxiv.org/abs/1703.05687>
- Saha, B., & Goebel, K. (2007). “Battery Data Set”, NASA Ames Prognostics Data Repository, Moffett Field, CA. Retrieved from <http://ti.arc.nasa.gov/project/prognostic-data-repository>
- Saha, B., & Goebel, K. (2009). Modeling Li-ion battery capacity depletion in a particle filtering framework. *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, 2909–2924. Retrieved from [https://www.phmsociety.org/sites/phmsociety.org/files/phm\\_submission/2009/phmc\\_09\\_38.pdf](https://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2009/phmc_09_38.pdf)
- Saha, B., Goebel, K., & Christophersen, J. (2009). Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the*

*Institute of Measurement and Control*, 31(3–4), 293–308. <https://doi.org/10.1177/0142331208092030>

- Sankararaman, S., Saxena, A., & Goebel, K. (2014). Are current prognostic performance evaluation practices sufficient and meaningful? *PHM 2014 - Proceedings of the Annual Conference of the Prognostics and Health Management Society 2014*.
- Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for Offline Evaluation of Prognostic Performance. *International Journal of Prognostics and Health Management*, (1), 1–20. Retrieved from [http://72.27.231.73/sites/phmsociety.org/files/phm\\_submission/2010/ijPHM\\_10\\_001.pdf](http://72.27.231.73/sites/phmsociety.org/files/phm_submission/2010/ijPHM_10_001.pdf)
- Si, X.-S., Zhang, Z.-X., & Hu, C.-H. (2016). *Data-Driven Remaining Useful Life Prognosis Techniques Stochastic Models, Methods and Applications*. China: Springer Series in Reliability Engineering. <https://doi.org/10.1007/978-3-662-54030-5>
- Tao, L., Cheng, Y., Lu, C., Su, Y., Chong, J., Jin, H., ... Noktehdan, A. (2017). Lithium-ion battery capacity fading dynamics modelling for formulation optimization: A stochastic approach to accelerate the design process. *Applied Energy*, 202, 138–152. <https://doi.org/10.1016/j.apenergy.2017.04.027>
- Wang, D., Yang, F., Zhao, Y., & Tsui, K.-L. (2017). Prognostics of Lithium-ion batteries based on state space modeling with heterogeneous noise variances. *Microelectronics Reliability*. <https://doi.org/10.1016/j.microrel.2017.06.002>
- Wu, L., Fu, X., & Guan, Y. (2016). Review of the Remaining Useful Life Prognostics of Vehicle Lithium-Ion Batteries Using Data-Driven Methodologies. *Applied Sciences*, 6(6), 166. <https://doi.org/10.3390/app6060166>
- Zhang, H., Miao, Q., Zhang, X., & Liu, Z. (2018). An improved unscented particle filter approach for lithium-ion battery remaining useful life prediction. *Microelectronics Reliability*, 81(24), 288–298. <https://doi.org/10.1016/j.microrel.2017.12.036>
- Zhang, J., & Lee, J. (2011). A review on prognostics and health monitoring of Li-ion battery. *Journal of Power Sources*, 196(15), 6007–6014. <https://doi.org/10.1016/j.jpowsour.2011.03.101>
- Zhang, X., Miao, Q., & Liu, Z. (2017). Remaining useful life prediction of lithium-ion battery using an improved UPF method based on MCMC. *Microelectronics Reliability*. <https://doi.org/10.1016/j.microrel.2017.02.012>
- Zhou, D., Yin, H., Fu, P., Song, X., Lu, W., Yuan, L., & Fu, Z. (2018). Prognostics for state of health of lithium-ion batteries based on gaussian process regression.

## BIOGRAPHIES



**Mikel Arrinda** received the B.S. degree in industrial electronic engineering in 2012 at MU, Mondragon, Basque Country, Spain. In 2013 completed his studies with a M.S. in integration of renewable energy sources into the electricity grid at EHU, Bilbao, Basque Country, Spain. After three years of activity in a private company, in 2017, he started his current research as a Research Scientist at CIDETEC institute for Energy Storage. His research is focused on applying various regression and state estimation techniques for predicting remaining useful life of energy storage systems based on lithium ion batteries, as well as developing a controllable and user friendly software for the sizing of the energy storage system of electric vehicles in terms of remaining useful life.



**Mikel Oyarbide** received the B.S. degree in automatic and industrial electronic engineering in 2009. Since then, he joined to IK4-CIDETEC working in the field of Battery management systems. In 2013 he obtained Phd degree in SoC and SoH estimator algorithms for Lithium ion, from the University of Mondragon. His current research interests are design of battery modules and packs and the integration of the systems in the real application.



**Haritz Macicior** received the B.S. degree in automatic and industrial electronic engineering at Mondragon University in 2002. In 2004 completed his studies with a M.Sc. degree in electric engineering at École Polytechnique de Montréal (EPM). Since 2005 he works in IK4-CIDETEC. Currently is the Head of Energy Storage Systems Unit. His present research interests are the development of battery storage systems for automotive and smart grids applications.



**Eñaut Muxika Olasagasti** is a lecturer and a researcher at Mondragon Unibertsitatea and he obtained his PhD in Electrical Engineering from the Institute National Polytechnique de Grenoble (INPG) in 2002. He has worked in machine-tool and power electronics control systems. His current research interests include reliability, availability, safety and performance modelling, model-based system engineering and adaptive hardware, software, communication system design.