Data-Driven State of Health and Functionality Estimation for Electric Vehicle Batteries Based on Partial Charge Health Indicators

Maite Etxandi-Santolaya, Tomas Montes, Lluc Canals Casals, Cristina Corchero, and Josh Eichman

***Abstract*—Effective Electric Vehicle (EV) operation relies on robust State of Health (SoH) estimation algorithms, key for informed battery management. Various algorithms have been proposed to estimate degradation, often depending on full charge segmentation or on conditions that deviate from real-world EV operation. Addressing this gap, this study introduces a cell-level SoH estimation algorithm based on partial charges. The proposed approach employs Health Indicators (HIs) derived from realistic laboratory testing, which contains a variety of voltage ranges during charge to replicate the complexity of real data. The study compares two commonly employed data-driven algorithms, Support Vector Regression (SVR) and Neural Networks (NN) and two estimation voltage ranges, which encompass the second and third Incremental Capacity (IC) peak. Along with the SoH, the battery functionality is estimated through the State of Function (SoF), leveraging degradation data and performance requirements for each tested cell. This enables the definition of an indicator quantifying the proximity of the battery to underperformance in specific applications. In general, the second IC peak shows higher correlation to the SoH. However, the NN SoH algorithm, when trained with high number of observations in the third IC peak, shows the lowest error with an average Root Mean Square Error (RMSE) of 0.00330. Moreover, the translation from SoH to SoF highlights the different performance requirements for each case and supports a functional definition of End of Life (EoL) beyond the fixed threshold.**

***Index Terms*—Electric vehicle, end of life, health indicator, machine learning, state of function, state of health.**

Received 11 March 2024; revised 16 August 2024; accepted 17 November 2024. Date of publication 25 November 2024; date of current version 18 April 2025. This work was supported in part by the European Union's Horizon 2020 research and innovation program under Grant 963580 (ALBATROSS project), in part by the assistance of the RYC2021-033477-I grant, funded by MCIN/AEI/10.13039/501100011033, in part by the European Union "NextGenerationEU"/PRTR, and in part by the Catalan Agency AGAUR under their research group support program grant to the ENMA research group number 2021 SGR 00226. This funding includes funds to support research work and open access publications. The review of this article was coordinated by Dr. Chongfeng Wei. *(Corresponding author: Maite Etxandi-Santolaya.)*

Maite Etxandi-Santolaya and Tomas Montes are with the Energy System Analytics, Catalonia Institute for Energy Research, 08930 Sant Adria de Besos, Spain, and also with the Department of Engineering Projects and Construction, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain (e-mail: metxandi@irec.cat).

Lluc Canals Casals is with the Department of Engineering Projects and Construction, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain.

Cristina Corchero is with the Energy System Analytics, Catalonia Institute for Energy Research, 08930 Sant Adria de Besos, Spain, and with the Bamboo Energy, 08018 Barcelona, Spain.

Josh Eichman is with the Energy System Analytics, Catalonia Institute for Energy Research, 08930 Sant Adria de Besos, Spain.

Digital Object Identifier 10.1109/TVT.2024.3505434

NOMENCLATURE

BMS	Battery management system.
BoL	Beginning of life.
CC	Constant current.
CV	Constant voltage.
DoD	Depth of discharge.
ECM	Equivalent circuit model.
EoL	End of life.
ETI	Equal time increase.
EV	Electric vehicle.
EVI	Equal voltage increase.
FEC	Full equivalent cycle.
HI	Health indicators.
IC	Incremental capacity.
ICA	Incremental capacity analysis.
IQR	Interquartile range.
IR	Internal resistance.
Li-ion	Lithium-ion.
ML	Machine learning.
MSE	Mean squared error.
NN	Neural networks.
RBF	Radial basis kernel.
RMSE	Root mean squared error.
RPT	Reference performance test.
SoC	State of charge.
SoF	State of function.
SoH	State of health.
SVR	Support vector regression.

I. INTRODUCTION

THE widespread adoption of the Electric Vehicle (EV) represents a step in the pursuit of a more sustainable transportation system. Central to the success of this transition is the efficiency and longevity of their batteries [1]. Among the numerous metrics used to assess the performance of EV batteries, the State of Health (SoH) measures the degradation, generally in terms of the capacity fade. The SoH provides insight to manufacturers, service providers, and consumers to make informed decisions regarding battery operation, maintenance, repair and replacement.

Various methodologies exist to estimate the SoH for EV batteries which can be categorized in physics-based, Equivalent Circuit Model (ECM) based, and data-driven approaches.

Physics-based models rely on fundamental electrochemical principles to estimate the SoH [2]. ECM-based methods combine the models with filtering techniques, like Kalman Filters, to track the model states [3]. These methods have important limitations stemming from high computational costs, especially in the case of physics-based algorithms, and lower accuracy for ECM-based methods [4].

Data-driven methods, in contrast, do not require battery models and utilize vast datasets from battery usage and performance to estimate the SoH. These methods have shown significant promise, benefiting from advanced techniques to discern complex patterns and correlations.

Several data-driven algorithms have been proposed for SoH estimation. Some algorithms utilize stress factor features, extracted from historical data, that define the degradation trend (e.g. temperature or Depth of Discharge (DoD)) [5], [6], [7], [8]. Implementing such methods requires higher storage capabilities and larger and complicated preprocessing due to data recording issues from lost connections, among others. Another alternative is to exclusively focus on battery response features for a given cycle. In this way, the SoH is estimated from the snapshot of battery response, that contains valuable information about the degradation.

In this context, one approach involves using the battery's operating curves for a particular cycle [9], [10], [11]. These curves include data on voltage, current, and temperature under specific conditions. However, the performance of the SoH algorithm can be improved by using more representative features, known as Health Indicators (HI) in the context of SoH estimation. These HIs are selected to capture essential information about the battery's condition and degradation. Leveraging these features within data-driven algorithms can lead to higher efficiency and accuracy.

A critical issue that affects these methods is the source of the training data. Many algorithms are trained using laboratory data, which often differs significantly from the dynamic and diverse conditions encountered by batteries in actual EVs. Laboratory data frequently assumes unrealistic scenarios, such as static operation, continuous identical driving cycles or full charges, which do not represent the complex and varied real-world operating conditions experienced by EV batteries. In line with the limitations of the laboratory data, several of the HIs proposed in the literature are based on unrealistic conditions, like constant discharge currents [12], [13], [14], [15], [16], [17].

Therefore, to be able to estimate the SoH from online data, the selection of HIs should be done considering their reproducibility in real-life. In this sense, the charging profiles offer more stable operating conditions than discharge. Thus, HIs obtained from charges hold higher interest and have been used to estimate the SoH [10], [11], [13], [14], [16], [18], [19], [20], [21]. Nevertheless, real-life charging profiles still show important variability, making it hard to find the same operating conditions over time. Among other differences, the charge does not always cover the same voltage range, as a consequence of changes in the driving cycles or a disconnection of the EV before reaching a full charge. For example, one study presented a SoH algorithm validated for real EVs [22], but assumed that the battery is discharged as much as the vehicle allows, which is not realistic for most users.

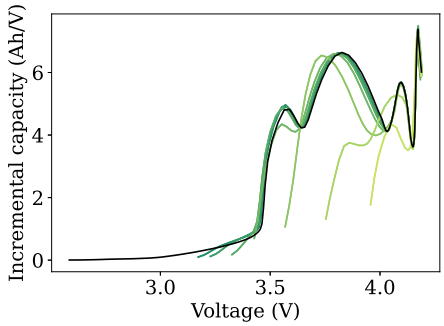


Fig. 1. IC curves of different partial charges obtained from laboratory experiments under the same conditions (C/3, 25 °C).

In contrast, a few studies have proposed partial charge based HI extraction. However, in many cases the HIs are obtained from dividing a full charge into segments [9], [23], [24], [25], [26]. This is generally done by selecting part of the CC charge, although one of the referenced studies focused on the CV part [26]. In another study the authors focus on part of a full multi-stage fast charge [27]. The approach taken by all these studies assumes that the curves of segmented full charges and partial charges are the same. Laboratory tests conducted on the same cell under the same charging current and temperature show that depending on the initial voltage, the extracted HIs change for the same SoH. Fig. 1 shows the Incremental Capacity (IC) curves of different partial charges and how they are affected by the initial voltage. This highlights an important issue that affects several of the reviewed SoH algorithms. Segmentation of full charges does not produce the same curves as partial charges and therefore, several of the HIs extracted are affected. Although the effect of the initial voltage has already been shown in the literature [28], several of the partial charge SoH estimation algorithms still employ charge segmentation.

To avoid this, others have focused on actual partial charges. One study proposed the use of the charge time from one State of Charge (SoC) level to another [29]. Similarly another study employed the height of the third IC peak and corrected the value based on the initial SoC of the charge [28]. Another study analysed real taxi data and selected specific charging sessions in a SoC range [30]. However, the accuracy of these methods is directly linked to the accuracy of the SoC algorithm. The voltage curves can be analysed to avoid errors that arise from using the SoC. Another study focused on the voltage curves, but always starting from the same voltage [31].

The literature review is summarized in Table I. It is important to emphasize that merely having an accurate SoH estimate may not be sufficient to make informed choices in managing EV batteries. Beyond understanding the current condition of the battery, it is equally crucial to predict when a battery might start exhibiting underperformance and eventually reach its End of Life (EoL).

TABLE I
SUMMARY OF THE LITERATURE REVIEW

1st Author	Ref	Year	Origin of the data		Type of feature		Source of features		
			Historical	Instantaneous	Full curves	HI extraction	Full charge	Constant discharge	Partial charge
A. Nuhic	[5]	2013	X						
G.-w. You	[6]	2016	X						
L. Song	[7]	2020	X						
Z. He	[8]	2021	X						
H. Huang	[9]	2022		X	X				X*
S. Khaleghi	[10]	2022		X	X		X		
Q. Gong	[11]	2022		X	X		X		
S. S. Sheikh	[12]	2020		X		X		X	
Z. Xia	[13]	2021		X		X	X	X	
W. Liu	[14]	2020		X		X	X	X	
W. Liu	[15]	2020		X		X		X	
X. Li	[16]	2023		X		X	X	X	
N. Yang	[17]	2022		X		X		X	
M. Lin	[18]	2023		X		X	X		
L. Cai	[19]	2020		X		X	X		
C. Zhang	[20]	2022		X		X	X		
M. Lin	[21]	2023		X		X	X		
A. Gismero	[22]	2023		X		X	X		
R. Xiong	[23]	2019		X		X			X
X. Feng	[24]	2019		X					X*
B. Gou	[25]	2021		X		X			X*
C. Zhang	[26]	2023		X		X			X*
Z. Zhou	[27]	2022		X		X			X*
B. Jiang	[28]	2020		X		X			X
E. Kheirkhah-rad	[29]	2023		X		X			X
R. Li	[30]	2022		X		X			X
R. Xiong	[31]	2023		X		X			X*

Note that the type of feature and source of curves HIs does not apply to historical data based studies. X represents partial charges derived from full charge segmentation.*

Conventionally, a SoH threshold of 70–80% has been considered as the point at which a battery is typically considered to reach the EoL for automotive purposes [32]. However, this approach does not account for the diverse driving requirements and patterns of individual users, which can significantly affect the practicality of this threshold.

To improve the accuracy of the EoL estimation, a driver-centric approach should be pursued through the introduction of the State of Function (SoF) [33]. The SoF measures the battery's functionality for a specific driver, considering the unique driving patterns and requirements of that individual. By understanding the interplay between battery degradation and performance, which is the aim of the SoF, a more tailored and effective strategy for battery EoL estimation can be established.

Considering the limitations of SoH algorithms highlighted, the first part of this study is focused on analysing laboratory data containing varied driving habits and charge voltage ranges. The data is first used to obtain representative HIs from full charges that give insight into the voltage ranges that contain essential information about the degradation. Based on this knowledge, a partial charge cell-level SoH algorithm is proposed based on two Machine Learning (ML) algorithms (Support Vector Regression (SVR) and Neural Networks (NN)). Then, in this study, the SoH is related to the SoF based on the characteristics of each case. The aspects that should be considered to define the SoF were presented in a previous study [33]. The current study puts into practice the estimation of the SoF based on battery data.

The main contributions of the study are listed below.

- Realistic dataset: analysis of a dynamic dataset obtained from laboratory testing, which more accurately reflects real-world driving habits, forming the basis for applicable SoH estimation methods by bridging the gap between laboratory and actual EV operation.
- Partial charge SoH estimation: the proposed algorithm leverages insights gained from full charge HI analysis and aligns with the dynamic nature of EV battery usage.
- Comparison of two popular ML algorithms (SVR and NN) in terms of computational cost and accuracy.
- Functionality estimation: this study goes a step beyond and considers the particularities of each case to define the SoF.

II. METHODOLOGY

A graphical representation of the methodology is presented in Fig. 2. This section starts by introducing the laboratory dataset in Section II-A, which contains: full charges where the SoH is measured and cycling that represent the EV driving, composed by driving cycles followed by partial charges.

The first part of the study is presented in Section II-B where HIs are extracted from full charges and their correlation with the SoH is analysed. The main result of this section is the determination of the voltage ranges most sensitive to degradation. Based on this output, Section II-C presents the SoH algorithm for partial charges that contain the target voltage ranges. Finally, in Section II-D, the degradation is combined with the EoL requirements derived from analysing the driving cycles to estimate the SoF.

A. Data Description

The data employed for this study contains the laboratory cycling of 6 cells labeled as ALB1, ALB2, ALB3, ALB4, ALB5 and ALB6. The cell model is the one being employed to build the battery packs for the EV in the H2020 Albatross project [34]. The cell is a Lithium-ion (Li-ion) cylindrical cell with NMC cathode from the manufacturer LG Chem, model INR21700M50LT. A summary of the cell characteristics is presented in Table II.

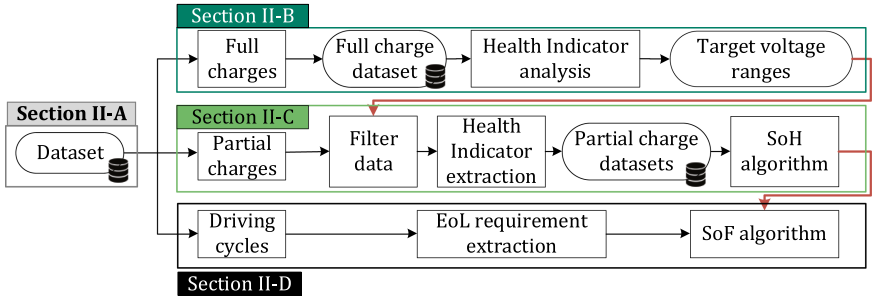


Fig. 2. Diagram showing the methodology of the study.

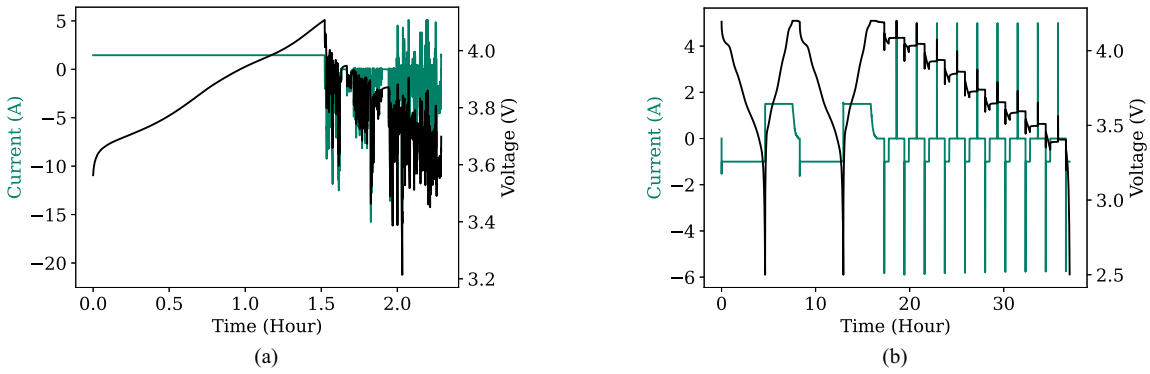
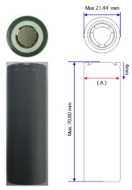


Fig. 3. Example voltage and current profiles of the dataset. (a) During a cycling test and (b) during a RPT.

TABLE II
CHARACTERISTICS OF THE CELL TESTED IN THE LABORATORY

Manufacturer	LG Chem
Model	INR21700-M50T
Positive Electrode	LiNiMnCoO2
Negative electrode	graphite and silicon
Diameter	21.44 mm
Length	70.80 mm
Weight	69.25 g
Nominal Capacity (Q_{nom})	4.85 Ah
Nominal Voltage	3.63 V
Charge cutoff Voltage	4.2 V
Discharge cutoff Voltage	2.5 V
Cutoff current	50 mA



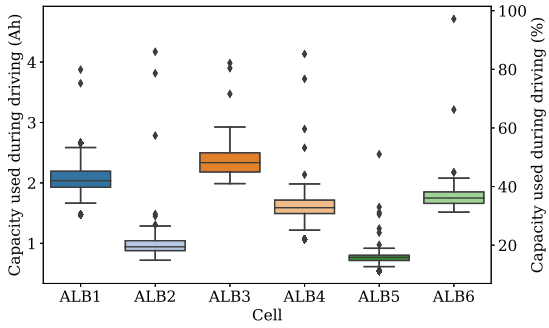


Fig. 4. Boxplots representing the capacity used during all historical driving cycles for each cell.

The testing protocol consists of cycling and Reference Performance Test (RPT)s. The cycling tests were performed at 35 °C, which corresponds to the maximum acceptable temperature permitted by the cooling system in the Albatross project. Thus, the temperature of 35 °C can be considered as part of normal operating conditions. The cycling test starts with a partial charge of the battery at C/3 until 4.05V and then a synthetic driving cycle, as shown in Fig. 3(a). Notice that negative currents represent discharge currents. The cycle is obtained from a synthetic driving cycle model [35].

Each of the cells tested aims to represent a driver that uses the EV daily to perform relatively homogeneous trips and charges the battery every day. Each synthetic cycle varies in duration, peak currents, resting times, etc. This aims to introduce variability common to real-world driving conditions, which are typically overlooked in standard cycling protocols. For the purpose of the study, it is important to highlight that the charge after each of the driving cycles covers different voltage range, since the DoD of each cycle is different.

Fig. 4 shows the boxplot of historical driving trips for each of the cells and Table III presents the main characteristics of the generated cycles for each cell, where Cd and Cr represent the C-rate during discharge and regenerative braking, respectively. Notice how cells ALB1 and ALB3 experience, on average, the deepest cycles. ALB1 and ALB4 record the highest average and peaks of discharge currents. These aspects will affect the driving requirements to be considered for the SoF estimation.

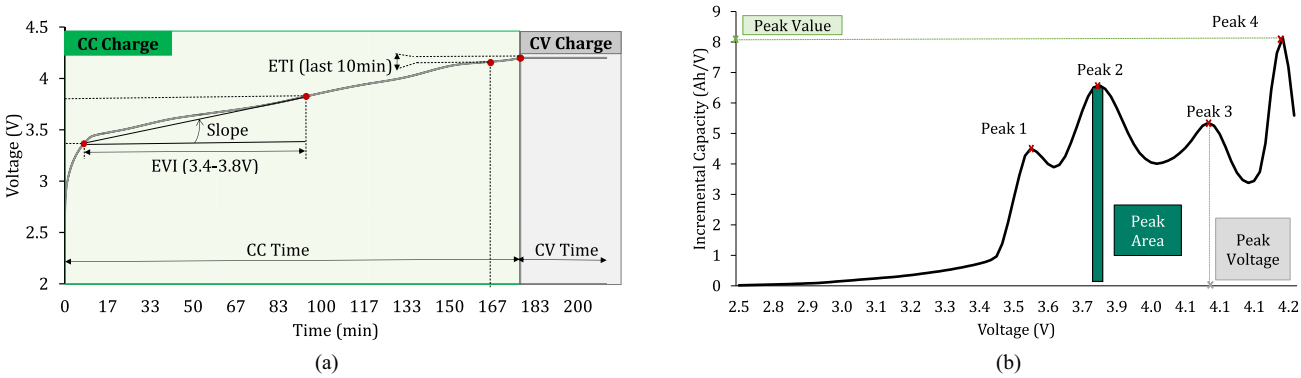


Fig. 5. Graphical representation of the full charge HIs for (a) the CC-CV curve and (b) the IC curve.

TABLE III
SUMMARY OF DRIVING CYCLE CHARACTERISTICS FOR EACH CELL

Cell	Av. Cd	Max. Cd	Av. Cr	Max. Cr
ALB1	0.73	4.23	0.27	2.63
ALB2	0.33	1.90	0.12	1.18
ALB3	0.34	1.90	0.12	1.18
ALB4	0.50	3.80	0.28	1.74
ALB5	0.22	1.71	0.12	0.78
ALB6	0.21	1.71	0.12	0.78

After approximately 50 Full Equivalent Cycle (FEC) a RPT is performed to measure the battery capacity and the Internal Resistance (IR) at different SoC values. The capacity test consists of two sets of a full Constant Current (CC) - Constant Voltage (CV) charge at C/3 and a full discharge at 1C. An example of a RPT is shown in Fig. 3(b). The SoH is obtained considering the average between both discharges, as shown in (1). The IR is measured after 1s and every 10% SoC using 1C current pulses. The battery is fully charged after the RPT.

$$SoH = \frac{Q_{dis1} + Q_{dis2}}{2 \cdot Q_{nom}} \tag{1}$$

where Q_{dis1} and Q_{dis2} are the discharged capacity obtained applying Coulomb Counting to each of the full discharges.

B. Health Indicators for Full Charges

A large variety of HIs have been proposed in the literature to estimate the SoH. In this study, the criteria employed for selecting the adequate ones considers the following points:

- Model-based HIs are not considered to adapt a fully data-driven approach.
- HIs that assume CC conditions during discharge are not considered due to their unrealistic nature.
- Differential Voltage Analysis based HIs are not consider as they contain similar information as Incremental Capacity Analysis (ICA) based ones, which are included.
- HIs defined based on the SoC are not considered to avoid the errors of the SoC estimation.
- HIs that cannot be obtained from the available data are not considered. This includes the voltage relaxation, as no rest period was considered after the driving cycles. Temperature related HIs are also not considered due to the use of the climate chamber.
- Since the IR affects the power capabilities, it is not considered as a HI considering that the SoH reflects the capacity or range capabilities instead.

TABLE IV
HEALTH INDICATORS EMPLOYED

Health Indicator		Label
CC Time		t_{CC}
CC Capacity		Ah_{CC}
CC Time Ratio		$t_{CC/total}$
CV Time		t_{CV}
CV Capacity		Ah_{CV}
Slopes at CC charge end	3.4-3.6 V	$SlopeCC_1$
	3.6-3.8 V	$SlopeCC_2$
	3.8-4 V	$SlopeCC_3$
	4-4.2 V	$SlopeCC_4$
Equal Voltage Increase (EVI)	2.6-3 V	EVI_1
	3-3.4 V	EVI_2
	3.4-3.8 V	EVI_3
	3.8-4.2 V	EVI_4
Equal Time Increase (ETI)	Last minute of CC charge	ETI_1
	Last 5 mins of CC charge	ETI_2
	Last 10 mins of CC charge	ETI_3
ICA	Peak 1 Voltage	IC_{V1}
	Peak 1 Value	IC_{P1}
	Peak 1 Area	IC_{A1}
	Peak 2 Voltage	IC_{V2}
	Peak 2 Value	IC_{P2}
	Peak 2 Area	IC_{A2}
	Peak 3 Voltage	IC_{V3}
	Peak 3 Value	IC_{P3}
	Peak 3 Area	IC_{A3}
	Peak 4 Voltage	IC_{V4}
	Peak 4 Value	IC_{P4}
	Peak 4 Area	IC_{A4}

Based on this criteria, Table IV summarizes the HIs obtained from the full charges of the RPTs and Fig. 5 represents their calculation based on the charge curves. The equations for the slope HIs, EVI, ETI are shown below. $Vmax$, $Vmin$ and dt depend on the range, as specified in Table IV.

$$Slope = \frac{Vmax - Vmin}{t_{Vmax} - t_{Vmin}} \tag{2}$$

$$EVI = t_{Vmax} - t_{Vmin} \tag{3}$$

$$ETI = V_{CCend} - V_{CCend-dt} \tag{4}$$

The IC curve is obtained from (5), with ΔV set to 15 mV. After testing different values, 15 mV effectively balanced noise reduction and the retention of key information. IC_V and IC_P are obtained by locating the peaks and the area is calculated using (6) which considers the adjacent points to the peak.

$$IC = \frac{dQ}{dV} \mid_k = \frac{Q_{k,2} - Q_{k,1}}{\Delta V} \tag{5}$$

$$IC_A = (V_{P+1} - V_{P-1}) \left(\frac{Q_{P+1} + Q_{P-1}}{2} \right) \tag{6}$$

To obtain the most representative HIs, the correlation between HIs and SoH is analysed. Unlike the Pearson correlation, which measures linear relationships, the Spearman correlation handles non-linear and monotonic relationships without such assumptions. Considering that the relationship between HI and SoH may not be strictly linear, in this study the Spearman correlation is employed.

C. Proposed State of Health Algorithm

From the analysis described in the previous section, the HIs that show the highest correlation to the SoH can be derived and the most relevant voltage ranges can be defined. The voltage range selection needs to balance important aspects such as reproducibility in real-life and sensitivity to degradation. Based on this, two models are built:

- P2: built with the voltage range that is most sensitive to degradation ($V_{t,P2}$). As will be shown, it corresponds to the second IC peak (3.6–3.9V).
- P3: to be able to estimate the SoH for cases where the partial charge voltage (V_c) does not fall into $V_{t,P2}$, a second model is built considering a different voltage range ($V_{t,P3}$). This last range shows less correlation to the SoH in full charges but contains a higher number of observations (3.9–4.05V).

As will be shown, the IC peak captured in the P2 model is the most sensitive to degradation. However, for users with a lower DoD, that do not reach the required voltage for the P2 model, the P3 model, which contains part of the third IC peak, can be used to obtain an estimation of the SoH. For cases where the charge voltage ranges contains both $V_{t,P2}$ and $V_{t,P3}$, the model with the highest accuracy will be prioritized.

The same methodology is applied to train both models, with the only difference being the voltage range considered for the features. The first step to build the SoH algorithm is to filter the charging sessions that encompass that target voltage range.

1) Partial Charge Dataset: Once the charging sessions that meet the criteria are filtered, the next step is to extract the new set of HIs obtainable during partial charges. The HIs selected are similar to those of the full charges but adapted to the partial voltage range. Different ranges were tested for the EVI and the best results were obtained with 3 divisions. 3 EVI features are obtained for the P2 model with 0.1V increments ($EVI1_{P2}$, $EVI2_{P2}$, $EVI3_{P2}$) and another 3 for the P3 model with 0.05V increments ($EVI1_{P3}$, $EVI2_{P3}$ and $EVI3_{P3}$).

TABLE V
FEATURE CALCULATION FOR EACH MODEL

Model	Feature	Calculation
P2	$EVI1_{P2}$	Eq. 3 for 3.6-3.7 V
	$EVI2_{P2}$	Eq. 3 for 3.7-3.8 V
	$EVI3_{P2}$	Eq. 3 for 3.8-3.9 V
	IC_{P2}	Max. of Eq. 5 in range 3.6-3.9 V
	IC_{A2}	Eq. 6 for P = IC_{P2}
	$minV$	Initial voltage of the charge
P3	$EVI1_{P3}$	Eq. 3 for 3.9-3.95 V
	$EVI2_{P3}$	Eq. 3 for 3.95-4.0 V
	$EVI3_{P3}$	Eq. 3 for 4.0-4.05 V
	IC_{P3}	Max. of Eq. 5 in range 3.9-4.05 V
	IC_{A3}	Eq. 6 for P = IC_{P3}
	$minV$	Initial voltage of the charge

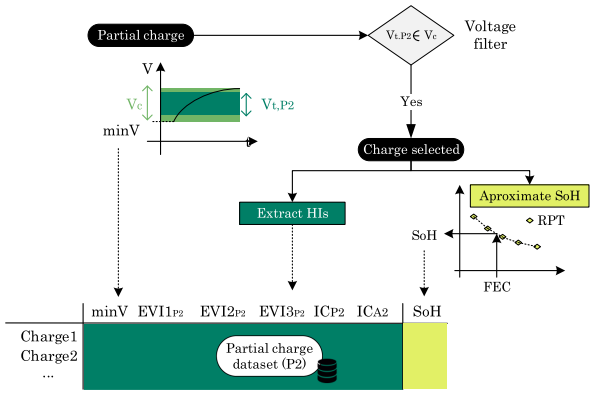


Fig. 6. Construction of the partial charge datasets. A partial charge corresponding to the P2 model is shown as an example.

Similarly, the peak value (IC_{P2} and IC_{P3}) and area below the peak (IC_{A2} and IC_{A3}) are obtained for the partial IC curves in both P2 and P3 target voltage ranges. The peak voltage is not considered as it showed lower correlation in the full HI analysis. Therefore, a set of 5 HIs are employed for each model.

As mentioned in the introduction, considering that the initial voltage of the charge can affect the extracted HIs (see Fig. 1), this voltage is included into the model inputs ($minV$). In this way, it is possible to avoid the errors derived from charge curve segmentation. The summary of features for the models and their calculation are summarized in Table V.

Since the capacity was only measured through RPTs, the SoH for each charge is approximated by averaging the SoH obtained in the previous and posterior RPTs, considering that the degradation is proportional to the FEC number. Even if the SoH is not a linear function of the FEC, this interpolation can be made since the FEC difference is not large.

A graphical representation of the process of building the partial charge dataset is provided in Fig. 6. 80% of the data is used for training and 20% for testing the model.

2) Algorithm Selection: Once the processed datasets are obtained, the next step is selecting the algorithm. In this study, two popular ML algorithms are selected and compared.

The first one is the SVR, which is a regression algorithm based on Support Vector Machines that aims to find a hyperplane that

best fits the training data while allowing for controlled deviations, known as the epsilon-insensitive tube. The tube represents a range around the predicted values within which no penalties are incurred.

The primal form of the objective function is given by (7). The function minimizes two terms: the square of the norm of the weight vector w, which encourages a small-margin hyperplane, and a penalty term for deviations within the epsilon-tube where ζ_i and ζ_i^* are slack variables representing errors. The regularization parameter C balances the trade-off between achieving a small training error and a simple model. The constraints ensure that the errors are within a specified margin (defined by ε) and non-negative. $\phi(x_i)$ represents the transformation of the feature vector x_i into a high dimensional space, allowing for the capture of nonlinear relationships.

$$\begin{aligned} \underset{w,b,\zeta,\zeta^*}{\text{minimize}} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\zeta_i + \zeta_i^*) \\ \text{subject to} \quad & y_i - w^T\phi(x_i) - b \le \varepsilon + \zeta_i \\ & w^T\phi(x_i) + b - y_i \le \varepsilon + \zeta_i^* \\ & \zeta_i, \zeta_i^* \ge 0 \end{aligned} \tag{7}$$

The optimization problem presented in (7) can be transformed into a dual optimization problem employing Lagrange multipliers. The detailed explanation of the dual formulation can be found in the literature [36]. The key advantage of the dual formulation is that it does not require that the transformed feature vector $\phi(x_i)$ is obtained and instead, only the inner products of the feature vectors needs to be computed. For this step, the kernel trick is employed which enables capturing nonlinear relationships in the data, without increasing the computational cost associated with feature transformation. In this study, the employed kernel is the Radial Basis Kernel (RBF), also known as the Gaussian kernel. The RBF measures the similarity between two input vectors in a high-dimensional space and is characterized by a parameter γ. The RBF kernel is given by (8) for input vectors x_i and x_j.

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \tag{8}$$

Given a set of support vectors (x_i, y_i), the predicted output for a new input x is given by (9).

$$f(x) = w^T\phi(x) + b = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi(x_i)^T\phi(x) + b \tag{9}$$

where α_i and α_i^* are the Lagrange multipliers corresponding to the Support Vectors.

The second employed ML model is the NN. A NN consists of interconnected layers of neurons, including an input layer, one or more hidden layers, and an output layer. Each neuron in a layer is connected to every neuron in the following layer through weighted connections. These weights determine the influence of the signal passed between neurons, enabling the propagation of information through the network.

The output of a neuron $z_j^{(l)}$ in layer l is determined by applying an activation function g to the weighted sum of its inputs plus a bias term, as shown in (10). Here, $w_{ij}^{(l-1)}$ represents the weight connecting neuron i in layer $l-1$ to neuron j in layer l, and $b_j^{(l)}$ is the bias term for neuron j in layer l. The activation functions commonly used in NNs include the sigmoid, tanh, and ReLU functions.

$$z_j^{(l)} = g\left(\sum_i w_{ij}^{(l-1)} z_i^{(l-1)} + b_j^{(l)}\right) \tag{10}$$

The training of a NN involves finding the optimal set of weights that minimize the error between the predicted outputs $\hat{y}$ and the actual outputs y. This is achieved by defining a loss function $L(y, \hat{y})$, such as the Mean Squared Error (MSE), as shown in (11), where n is the number of samples.

$$L(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{11}$$

To minimize the loss function, optimization algorithms like the Adam optimizer are employed. The gradients of the loss function with respect to the weights are computed using backpropagation. The weight update rule during training is given by (12), where η is the learning rate, $\frac{\partial L}{\partial w_{ij}^{(l)}}$ represents the gradient of the loss function with respect to the weight w_{ij}, and w_{ij} is the weight to be updated.

$$w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta\frac{\partial L}{\partial w_{ij}^{(l)}} \tag{12}$$

The training process involves iteratively updating the weights across all layers to reduce the error.

Hyperparameter tuning is an important aspect of both SVR and the NN. In the case of SVR, hyperparameters include the choice of kernel (and value of γ in the case of RBF), the regularization parameter C and the ε-insensitive tube width. For the NN, hyperparameters include the architecture of the neural network (number of layers and neurons) and the batch size during training. For both models, grid search with cross-validation is employed to tune the hyperparameters. The cross-validation procedure involves splitting the training dataset into five folds, systematically training the model on four folds while validating on the remaining one. The objective is to minimize the MSE during this process. The grid search explores various combinations of the hyperparameters. For SVR the hyperparameters tested are C (0.1, 0.5, 1, 10), ε (0.01, 0.1, 0.2, 0.5), and γ (0.001, 0.01, 0.1, 1, 10). For the NN, different numbers of hidden layers (1, 2, 3), neurons per layer (10, 50, 100) and batch sizes (16, 32, 64) are included, while the learning rate and epoch number are set to 0.001 and 50, respectively. The resulting best combination of hyperparameters is determined based on the configuration that yields the lowest MSE across the cross-validated folds.

The models are trained using Python's TensorFlow 2.14 and scikit-learn 1.2.2 f with a computer with the following specifications: Intel Core i5, 4 cores, 16GB RAM.

D. State of Function Algorithm

The SoF is used in this study to quantify the functionality of the battery at a given state for a specific application. Thus,

the SoF indicates how far the battery is from underperforming and reaching the EoL, where the SoF drops to 0%. Unlike the common approach, in this study, the EoL is derived for each application based on analysing the performance requirements and used to calculate the SoF.

Two main events may force the EoL: capacity fade which is related to the SoH, and the power fade which is related to the IR increase. Therefore, these two aspects are included in the SoF. As shown by (13), the most restrictive of both terms will define the functionality of the battery.

$$SoF = \min\left(SoF_c, SoF_{IR}\right) \tag{13}$$

1) Capacity Requirements (SoF_c): The SoF_c measures how far the current capacity (C) is from the EoL capacity threshold (C_{EoL}). To define C_{EoL} the historical range requirements of the driver are analysed. The approach considered in this study is to discard the historical trips that do not represent the normal behaviour of the driver (outliers). These extreme trips are assumed to be covered with alternative transportation methods or by considering intermediate fast-charging. To obtain the C_{EoL}, all the historical trip consumptions are analysed and a 1.5 multiplication factor is applied to the Interquartile Range (IQR), which represents the upper whisker in the boxplot (Fig. 4). Thus, any value above this C_{EoL} is considered to be an outlier. Equation (14) shows the calculation of the SoF_c where C_{BoL} is capacity at Beginning of Life (BoL). C_{BoL} is obtained from the first capacity measurement.

$$SoF_c = \frac{C - C_{EoL}}{C_{BoL} - C_{EoL}} \tag{14}$$

2) Power Requirements (SoF_{IR}): For a given driving cycle, the IR plays an important role in the ability to perform the driving trip appropriately. Following Ohm's law, the voltage in the battery after a discharge current drops proportionally to the value of the IR. Thus, if the IR surpasses a critical threshold, the voltage would drop below the minimum acceptable level. One of the tasks of the Battery Management System (BMS) is to limit the discharge power to avoid this situation.

Therefore, to define the SoF_{IR} it is necessary to understand power constraints that emerge during driving. These constraints are notably dependent on individual driving habits, with abrupt accelerations leading to spikes in current and higher DoDs resulting in lower voltages. Consequently, a detailed analysis of the driving profiles is crucial.

To define the SoF_{IR}, the chosen approach involves estimating, for each timestep i of each historical driving cycle, the IR (IR_i) and its maximum permissible value ($IR_{i,\max}$). The calculation of IR_i involves (15). Similarly, $IR_{i,\max}$ is determined by (16) where, instead of U_i the minimum acceptable voltage of the battery ($U_{\min}$) is considered. Therefore, $IR_{i,\max}$ represents the maximum value of the IR before the undervoltage appears. The time between measurements i and $i+1$ is set to 1s.

$$IR_i = \frac{U_i - U_{i-1}}{I_i - I_{i-1}} \tag{15}$$

$$IR_{i,\max} = \frac{U_{\min} - U_{i-1}}{I_i - I_{i-1}} \tag{16}$$

For each driving cycle, the most critical point is identified, where IR_i is closest to ($IR_{i,\max}$). This point is then utilized to define the $SoF_{IR,cycle_n}$ for that particular cycle n, as indicated by (17). IR_{BoL} represents the IR at the BoL (measured during the first cycling) and IR_n and IR_{max_n} the values of the IR and maximum IR for the critical point. $SoF_{IR,cycle_n}$ is not allowed to exceed 1.

It is essential to note that the $SoF_{IR,cycle_n}$ may vary in subsequent cycles where the power requirements are different. To account for this variability, all driving cycles between RPTs are considered, and the SoF_{IR} for this entire period is determined by selecting the lowest recorded value ((18) where N represents the number of driving cycles between RPTs). This allows to define the SoF_{IR} not from a single cycle but considering the historical power requirements of the driver. In a real BMS instead of the driving cycles between RPTs a selection based on a time span can be made (e.g. one month).

$$SoF_{IR,cycle_n} = \min\left(1, \frac{IR_{\max_n} - IR_n}{IR_{\max_n} - IR_{BoL}}\right) \tag{17}$$

$$SoF_{IR} = \min_{n=1}^{N}\left\{SoF_{IR,cycle_n}\right\} \tag{18}$$

III. Results

The evolution of the SoH over FEC and over time is represented in Figs. 7(a) and (b) respectively. Since the cells have previously undergone calendar ageing and some unrecorded cycling, the initial SoH is not 1 and between 0.93 and 0.94 for all cells. Depending on the cell, the rate of degradation changes: ALB5 undergoes the fastest degradation over FEC and ALB3 the slowest one. A similar tendency can be found when looking at the evolution of SoH over time, even though the trends are more similar. Understanding the causes of the degradation tendencies is beyond the scope of this work and a different study will explore the impact of the different stress factors, such as current peaks and driving DoD, in detail.

Along with the capacity, during the RPTs, the IR was derived. As an example, Fig. 7(c) shows the evolution of the IR at 100% SoC over FEC. Although some variability is observed, the IR shows an increasing trend over FEC, as expected.

A. Full Charge Health Indicator Analysis

First, the HIs corresponding to the full charges are presented. A sub-selection of HIs with a Spearman correlation greater than 0.8 is then obtained, which are, in order of correlation, Ah_{CV}, IC_{P3}, IC_{P2}, $SlopeCC_2$, Ah_{CC}, t_{CC}, $SlopeCC_4$, EVI_4, EVI_3, and IC_{A2}. Since Ah_{CC} is directly proportional to t_{CC}, it is no longer considered. Table VI shows the Spearman correlation values between the HIs and the SoH.

Fig. 8 represents the CC-CV and IC curves for various degradation levels, including zoomed-in views of the voltage ranges where the second and third IC peaks take place. Note that visualizing the degradation trend in the CC-CV curve is not straightforward. In fact, the IC curve is introduced to help identify movements more clearly. As the Spearman correlation confirms, the second and third peak value in the IC curve exhibits

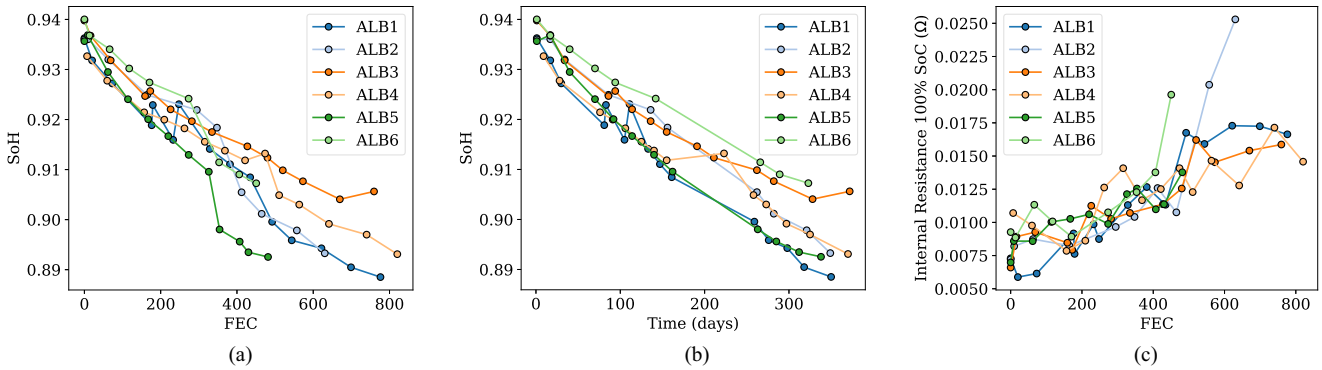


Fig. 7. Evolution of the degradation for each cell. (a) SoH over FEC, (b) SoH over time and (c) IR at 100% SoC over FEC.

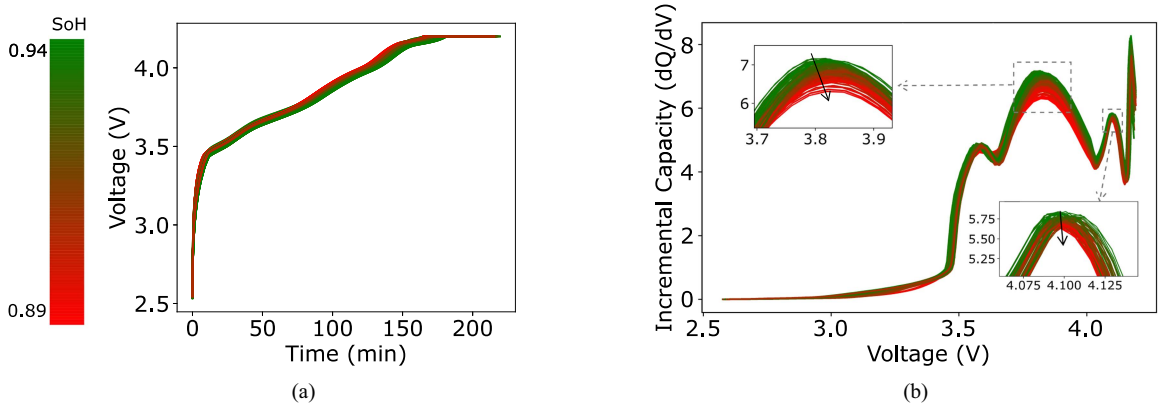


Fig. 8. Evolution of the charge curves over time. (a) Voltage over time, (b) IC curve with zoom-in views.

TABLE VI
SPEARMAN CORRELATION FOR FULL CHARGE BASED HIS SELECTED

HI	Spearman Correlation
Ah_{CV}	0.992
IC_{P3}	0.965
IC_{P2}	0.952
$SlopeCC_2$	-0.928
t_{CC}	0.926
$SlopeCC_4$	-0.898
$EVI4$	0.884
$EVI3$	0.883
IC_{A2}	0.858

a clear tendency with degradation. In particular, the second IC peak moves down and slightly to the right, while the area below the peak decreases. The third peak also shifts to lower values but the horizontal movement is not so clear and neither is the change in the area below the peak. The slopes $SlopeCC_2$ and $SlopeCC_4$ tend to increase with lower SoH values, which aligns with the increase of EVI_3 and EVI_4. Intuitively, t_{CC} and Ah_{CV} decrease with degradation.

These results are leveraged to find the voltage range that is considered for the partial charge SoH estimation. Selecting an adequate voltage range involves a delicate balance between sensitivity to degradation and adherence to practical constraints set by battery manufacturers and common charging conditions.

Considering the results of this section and the associated voltage ranges for each HI, a compelling choice may lie in a mid-charge voltage window. The second IC peak occurs in the voltage range 3.7-3.9V, which is sensitive to degradation as supported by the high correlation of IC_{P2}, IC_{A2}, $SlopeCC_2$, $EVI3$ and $EVI4$. Another sensitive voltage range would be the one corresponding to the third IC peak. However, this range is close to the maximum voltage of the battery which may be restricted by the manufacturer to avoid overvoltage. Therefore, the voltage range selected for the P2 model corresponds to the second IC peak and extended to include more observations ($V_{t,P2} = 3.6$–3.9V). In this range, degradation effects are likely to manifest and represent a mid-SoC charge of the battery which aligns with normal EV charging patterns.

As mentioned in the methodology, a user with low DoDs during driving may not reach the minimum voltage of 3.6V often and for those cases a higher voltage range is selected ($V_{t,P3} = 3.9$–4.05V) to train a second model.

B. State of Health Estimation Based on Partial Charges

Based on the derived target voltages, the partial charge datasets are obtained, as explained in Section II-C and the tuned SVR and NN models are built. Table VII provides an overview of the distribution of charges among both models, revealing the number of charges attributed to each. Notably, cells ALB2,

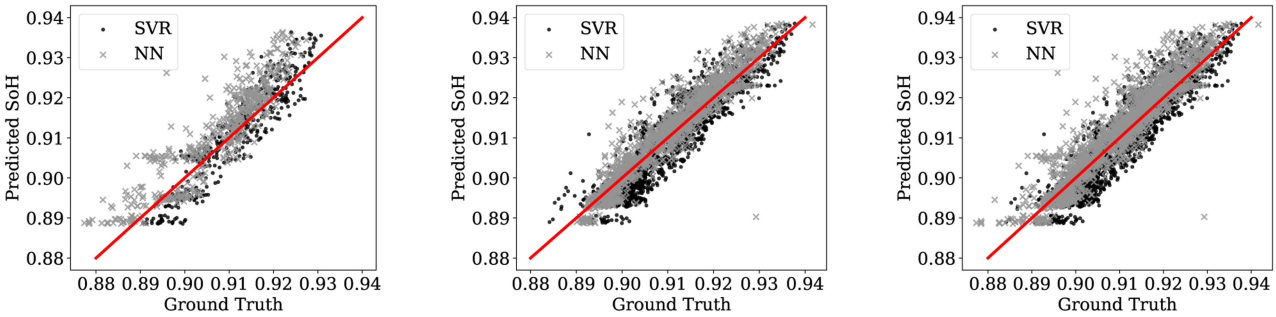


Fig. 9. SoH predictions over ground truth for (a) P2, (b) P3 and (c) full model. *The closer the points are to the 45° red line, the more accurate the predictions.*

TABLE VII
NUMBER OF CHARGES IN THE P2 AND P3 MODELS

Cell	Total charges	P2	P3	P2+P3
ALB1	1724	63.8%	98.0%	63.7%
ALB2	3213	0.5%	66.1%	0.5%
ALB3	1421	56.7%	96.1%	54.1%
ALB4	2455	13.4%	81.2%	7.5%
ALB5	3084	0.0%	59.9%	0.0%
ALB6	1294	6.0%	61.6%	0.3%
Total	**13191**	**17.7%**	**74.4%**	**15.7%**

ALB5 and ALB6 exhibit minimal representation within the P2 model's voltage range, and therefore rely on the P3 model for estimating the SoH. This observation underscores the distinctive charging characteristics of these cells caused by low DoDs. Notice how almost all charges that fall into the P2 voltage range also fall into the P3 one, and only in a limited number of cases the voltage stopped before 4.05V.

Table VIII shows the hyperparameters of the tuned models, their associated Root Mean Squared Error (RMSE) and other error metrics on the test datasets after the grid search with cross-validation, along with inference and training time. Note that inference time refers to the time to obtain an estimation of the SoH based on a given set of HIs. The extraction of HIs given a charging curve takes approximately 0.09 s. The SVR has the lowest training and inference time. For the NN built, they are higher but do not pose a constraint for the application.

For the SVR algorithm, the P2 model exhibits slightly reduced errors and higher value of R^2 compared to the P3 model, although the difference is not remarkable. The tuned hyperparameters show that the C value is larger for the P2 model and thus, a higher penalty is imposed on deviations from the actual data points during training, resulting in more complex decision boundaries. On the other hand, the smaller C for the P3 model may indicate a relatively higher tolerance for errors in fitting, allowing it to generalize better to a larger set of observations.

On the other hand, for the NN, results show that P3 has lower errors than P2. Even if some observations in P3 do not encompass the second IC peak, which is more sensitive to degradation, the high number of observations makes the NN accuracy increase. In fact, the P3 NN model is the most accurate one out of the four models built. This shows that if a high number of datapoints is available, the NN can provide accurate estimations even for higher voltage ranges.

Fig. 9 shows all the estimated values of the SoH compared to the ground truth for the P2, P3 and full models. Notice that the term full is employed to describe the result of combining both estimations, but prioritizing the the most accurate when common partial charges are observed (P2 for SVR and P3 for NN). This model produces the highest number of estimations.

Fig. 10 shows the RMSE for each cell and all models built. The P2 SVR model shows RMSE values below 0.005 for all cells except for ALB2, which presents the highest error (0.00761). This is likely due to the limited observations in the test dataset, which can result in higher variability in the evaluation metrics. The lowest error in the P2 SVR model is obtained for ALB3 with a RMSE of 0.00403. For the P2 NN model, the errors are higher for all cells (reaching up to 0.0093 for ALB6), except for ALB2 which is the only case in P2 where the NN outperforms the SVR.

The P3 SVR model errors tend to be slightly higher, up to 0.00575 for ALB1, compared to the P2 SVR model. Nevertheless, the lowest error for the SVR models is found for ALB3 with an RMSE of 0.00396. As discussed, the P3 NN model is the most accurate one (reaching a minimum error of 0.0027 for ALB2) except for one of the cells (ALB1) where the P2 SVR model shows lowest error.

Considering the full model, the average RMSE for all cells is 0.00468 and 0.00434 for SVR and NN respectively. For cells with high observations in P2, the errors tend to increase in the full model, considering the NN case due to its limited accuracy (e.g. ALB1). The opposite is observed when considering the SVR model.

For the final part of the study, the most accurate model is selected (P3 NN). To enhance the stability and reliability of the SoH predictions, an averaging approach is employed over a specified window of 10 FECs. This entails computing the average SoH prediction for all estimations made within this window. Consequently, SoH estimations performed for charging events with fewer than 10 FECs between them are averaged. Fig. 11 illustrates the averaged approach, where the SoH measured during the RPTs is juxtaposed with the averaged SoH predictions over the FEC window. Through this approach, it

TABLE VIII
TUNED HYPERPARAMETERS FOR THE MODEL AND ERRORS

Model		Hyperparameters	$RMSE$	MAE	MSE	R^2	Inference Time	Training Time
P2	SVR	C = 10, ε = 0.01, γ = 0.01	0.00463	0.00385	2.15E-5	0.854	0.02 ms	0.3 s
	NN	Layers = 2, Neurons = 64, Batch = 16	0.00668	0.00528	4.46E-5	0.697	0.6 ms	35 s
P3	SVR	C = 0.5, ε = 0.01, γ = 0.01	0.00472	0.00385	2.22E-5	0.836	0.05 ms	1.6 s
	NN	Layers = 2, Neurons = 64, Batch = 16	0.00356	0.00249	1.27E-5	0.906	0.3 ms	141 s

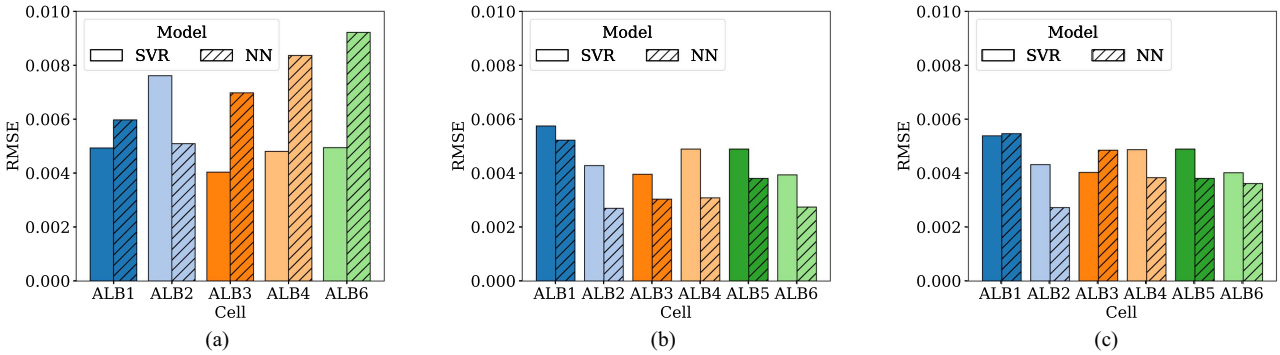


Fig. 10. RMSE for each cell considering models (a) P2, (b) P3 and (c) full. *Note that ALB5 does not have any predictions for P2.*

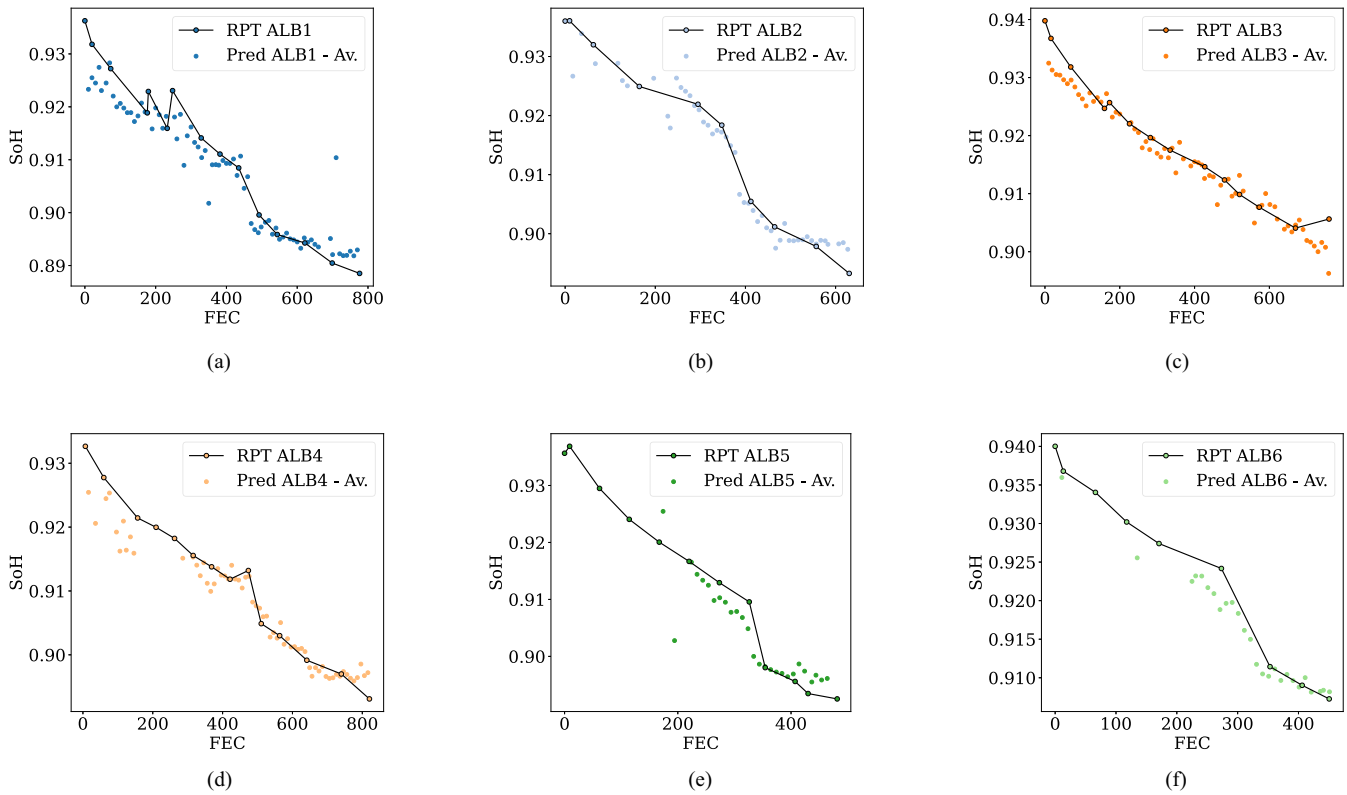


Fig. 11. Averaged SoH predictions in a range of 10 FEC over FEC compared with the RPT measurements for (a) ALB1, (b) ALB2, (c) ALB3, (d) ALB4, (e) ALB5 and (f) ALB6.

is possible to mitigate the potential impact of variability in individual predictions and provide a smoothed representation of the cell's health status over consecutive charges. Notice that, additionally, a restriction on the time between estimations can be added to make sure that the calendar ageing effects are not discarded. However, for the employed data this is not a concern as no resting time between cycles was considered. With this averaged approach, the model RMSE is found to be 0.00330. The RMSE for each cell is presented in Fig. 12 which ranges from 0.00230 (ALB6) to 0.00440 (ALB1).

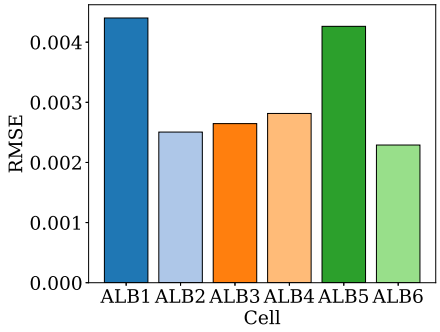


Fig. 12. RMSE for each cell considering the averaged model.

TABLE IX
EoL Capacity Threshold for Each Cell Based on Historical Requirements

Cell	EoL Ah	EoL SoH
ALB1	2.589	51.8%
ALB2	1.288	25.8%
ALB3	2.968	59.4%
ALB4	2.043	40.9%
ALB5	0.928	18.6%
ALB6	2.131	42.6%

C. State of Function

After the estimation of the SoH, the individual driving requirements are considered to estimate the SoF. The first part of this estimation is related to the capacity requirements. Considering the capacity required to meet each of the historical cycles of a driver it is possible to define a threshold on the capacity. By discarding the outliers (see Fig. 4), the EoL threshold for the capacity is defined in the upper whisker, as shown in Table IX. Considering these values, the cells with most demanding capacity requirements are ALB1 and ALB3. Considering these values, the evolution of SoF_c is presented in Fig. 13(a).

The second part of the SoF is related to the IR. The SoF_{IR} is presented in Fig. 13(b). Notice how the SoF_{IR} shows slightly similar trends as the IR measured in the RPTs (Fig. 7(c)). For example, at the last cycles of ALB2 an increase in IR was observed, which causes the SoF_{IR} to decrease.

When combining both SoF_c and SoF_{IR}, the general functionality (SoF) of the battery is obtained, as presented in Fig. 13(c). It can be observed that for some cells, the most restrictive constraint is the capacity one. For example, for both ALB3 and ALB5 the SoF follows the same trend as SoF_c. This shows that under the health and operational conditions tested, power capabilities do not restrict the performance of those cells. In other cases, both capacity and power constraints are observed, meaning that the SoF at points corresponds to SoF_c and at others to SoF_{IR}. For example, in ALB2 power constraints are observed in the beginning which translate into capacity ones eventually. Finally, due to the IR increase at the end of the cycling of ALB2, the last estimation of SoF shows that the power is more restrictive at that point.

At this point, it is important to highlight how the individualities of each cell, both in driving requirements and degradation trends (capacity and IR) affect the SoF. This emphasizes the importance of understanding battery functionality along with the degradation, which is overlooked in the current literature and can be done through the SoF proposed in this study.

IV. Discussion

The proposed SoH algorithm adopts a more practical approach by avoiding the need for full charges, relying instead on partial charges, which align with real-world scenarios. Notably, compared to the existing studies, the presented methodology avoids compounding errors from the SoC estimation and, most importantly, it avoids the inaccuracies derived from charge segmentation proposed in many of the existing studies. The results obtained from the SoH algorithm demonstrate errors within the range reported in the reviewed literature studies (Table I), which range from 0.00027 to 0.04 (RMSE), compared to 0.00330 from the averaged NN model in this study.

The boundaries of the SoH estimation are defined by the available laboratory testing conditions. On account of the available data, the SoH range that the algorithm can predict is 0.94-0.89. This range will be extended in the future when more data is available, in order to evaluate the model accuracy for lower SoH values. Regarding the operating conditions, a temperature of 35 °C and a charging current of 1.5 A is required for the estimation. This implies that the SoH estimation is performed when the charge occurs at high temperatures (e.g. right after a driving trip when the battery has self-heated up) and employing slow charging, which seems to dominate the charging trends according to the IEA report [37].

To estimate the SoH based on other temperatures and charging currents, three methodological frameworks can be considered. The first option (Option A), is to build a reduced number of algorithms at specific conditions. The same C-rate can be employed, assuming that the user charges the battery at similar powers considering the nominal powers for private level 2 AC charging (3.7, 7.4, 11 and 22 kW). Different temperatures can be considered to reflect the seasonal changes over the year. With this Option A, the SoH is only estimated once the average temperature during the charge and the C-rate corresponds to one of the trained models.

A more advanced alternative (Option B) requires the use of Transfer Learning to, starting with the baseline model, estimate the SoH once new data at different operation conditions is available.

The final option (Option C) includes the average temperature and C-rate as features of the model. In this case, the SoH would be estimated every time that a charge is recorded, as long as it covers the target voltage range. Therefore, this option allows for a more frequent SoH estimation, but requires a large number of observations.

As described in the testing protocol (Section II-A), the same C-rate and temperature are considered for all the charges. This made it impractical to employ Option C for the dataset and, for that reason, the selected approach was to build a single model which is the first step for both Option A and Option B.

Thus, this study, starting at the cell level and under specific charging conditions, establishes a foundational understanding necessary for real-life SoH estimations at pack level. Future

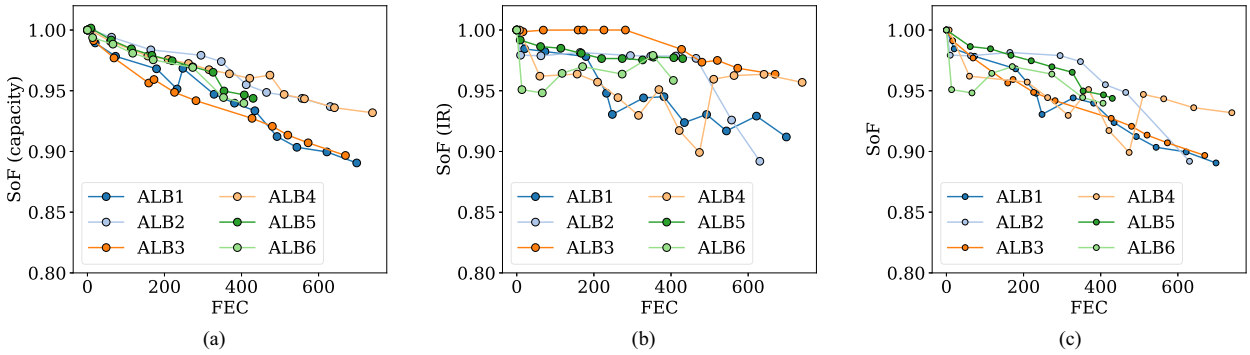


Fig. 13. State of Function for each cell over FEC (a) Capacity SoF, (b) IR SoF, and (c) total SoF.

work includes the crucial step of implementing the presented methodology on board an EV, where real data collected from the BMS will replace laboratory-based use cases. To do so, several challenges will be addressed, including the analysis of different chemistries, the onboard estimation of the IR, cell to cell ageing variability in the battery pack, the impact of cell balancing and thermal management, and the influence of varying ambient temperatures and C-rates during charging. In addition, forecasting techniques will be explored for the SoF based on the operation stress factors in order to estimate the EoL without assuming the fixed threshold, resulting in an individualized and improved estimation.

V. Conclusion

This paper presents a SoH estimation algorithm based on partial charges, leveraging a realistic dataset obtained from laboratory testing that reflects real-world driving habits. The algorithm breaks away from the limitations of existing methods by employing only BMS measurements and avoiding charge segmentation, recognizing the inherent differences between the curves of segmented full charges and partial charges. The presented approach, at cell-level, represents the first step to construct pack-level SoH estimations that bridge the gap between laboratory conditions and actual EV operation.

The study has compared two estimation voltage ranges: one close to the second IC peak (P2), shown to be most correlated to the SoH, and another close to the third (P3). In addition, two common Machine Learning models have been compared (SVR vs NN).

The NN is very sensitive to the number of datapoints, and produced less precise estimations for P2 than the SVR, achieving a RMSE of 0.00668 compared to 0.00463. However, since for P3 more observations are available, the NN outperformed the rest of the models with an average RMSE of 0.00356. The P3 NN model is further enhanced by averaging over a window of FECs, resulting in smoother predictions and a lower RMSE of 0.00330.

Along with the SoH, this study has also estimated the battery functionality through an improved definition of the SoF. The SoF is composed of two parts, one related to the capacity loss and another to the power loss. Depending on the ageing trends and driving requirements, for each cell either capacity or power constraints have shown to mark the cell functionality, showcasing the need to include both aspects in the SoF definition.

Results show how, even if the SoH is similar, the SoF highly depends on the usage and type of the battery. This highlights the importance of analysing battery functionality along with the SoH which allows for a more complete understanding of the requirements of each application. In this way, the EoL methodologies are improved by analysing individual requirements instead of assuming the universal SoH threshold of 70–80%.

References

[1] G. Haddadian, M. Khodayar, and M. Shahidehpour, "Accelerating the global adoption of electric vehicles: Barriers and drivers," *Electricity J.*, vol. 28, no. 10, pp. 53–68, Dec. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S104061901500250X

[2] J. Li, K. Adewuyi, N. Lotfi, R. Landers, and J. Park, "A single particle model with chemical/mechanical degradation physics for lithium ion battery state of health (SOH) estimation," *Appl. Energy*, vol. 212, pp. 1178–1190, Feb. 2018.

[3] G. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs Part 1. background," *J. Power Sources*, Jun., vol. 134, no. 2, pp. 252–261, 2004.

[4] M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," *Nat. Mach. Intell.*, vol. 2, no. 3, pp. 161–170, Mar. 2020.

[5] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, "Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods," *J. Power Sources*, vol. 239, pp. 680–688, Oct. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378775312018605

[6] G.-w. You, S. Park, and D. Oh, "Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach," *Appl. Energy*, vol. 176, pp. 92–103, Aug. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261916306456

[7] L. Song, K. Zhang, T. Liang, X. Han, and Y. Zhang, "Intelligent state of health estimation for lithium-ion battery pack based on Big Data analysis," *J. Energy Storage*, vol. 32, Dec. 2020, Art. no. 101836. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352152X2031673X

[8] Z. He, X. Shen, Y. Sun, S. Zhao, B. Fan, and C. Pan, "State-of-health estimation based on real data of electric vehicles concerning user behavior," *J. Energy Storage*, vol. 41, Sep. 2021, Art. no. 102867. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352152X21005892

[9] H. Huang et al., "An enhanced data-driven model for lithium-ion battery state-of-health estimation with optimized features and prior knowledge," *Automot. Innov.*, vol. 5, no. 2, pp. 134–145, Apr. 2022.

[10] S. Khaleghi et al., "Developing an online data-driven approach for prognostics and health management of lithium-ion batteries," *Appl. Energy*, vol. 308, Feb. 2022, Art. no. 118348.

[11] Q. Gong, P. Wang, and Z. Cheng, "A data-driven model framework based on deep learning for estimating the states of lithium-ion batteries," *J. Electrochem. Soc.*, vol. 169, no. 3, Art. no. 030532, Mar. 2022.

[12] S. S. Sheikh et al., "A battery health monitoring method using machine learning: A data-driven approach," *Energies*, vol. 13, no. 14, Jul. 2020, Art. no. 3658. [Online]. Available: https://www.mdpi.com/1996-1073/13/14/3658

[13] Z. Xia and J. A. A. Qahouq, "Lithium-ion battery ageing behavior pattern characterization and state-of-health estimation using data-driven method," *IEEE Access*, vol. 9, pp. 98287–98304, 2021.
[14] W. Liu, Y. Xu, and X. Feng, "A hierarchical and flexible data-driven method for online state-of-health estimation of li-ion battery," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14739–14748, Dec. 2020.
[15] W. Liu and Y. Xu, "Data-driven online health estimation of li-ion batteries using a novel energy-based health indicator," *IEEE Trans. Energy Convers.*, vol. 35, no. 3, pp. 1715–1718, Sep. 2020.
[16] X. Li, L. Ju, G. Geng, and Q. Jiang, "Data-driven state-of-health estimation for lithium-ion battery based on aging features," *Energy*, vol. 274, Jul. 2023, Art. no. 127378.
[17] N. Yang, Z. Song, H. Hofmann, and J. Sun, "Robust state of health estimation of lithium-ion batteries using convolutional neural network and random forest," *J. Energy Storage*, vol. 48, Apr. 2022, Art. no. 103857. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352152X21015231
[18] M. Lin, Y. You, J. Meng, W. Wang, J. Wu, and D.-I. Stroe, "Lithium-ion batteries SOH estimation with multimodal multilinear feature fusion," *IEEE Trans. Energy Convers.*, vol. 38, no. 4, pp. 2959–2968, Dec. 2023.
[19] L. Cai, J. Meng, D.-I. Stroe, J. Peng, G. Luo, and R. Teodorescu, "Multi-objective optimization of data-driven model for lithium-ion battery SOH estimation with short-term feature," *IEEE Trans. Power Electron.*, vol. 35, no. 11, pp. 11855–11864, Nov. 2020.
[20] C. Zhang, S. Zhao, Z. Yang, and Y. Chen, "A reliable data-driven state-of-health estimation model for lithium-ion batteries in electric vehicles," *Front. Energy Res.*, vol. 10, Sep. 2022, Art. no. 1013800.
[21] M. Lin, C. Yan, W. Wang, G. Dong, J. Meng, and J. Wu, "A data-driven approach for estimating state-of-health of lithium-ion batteries considering internal resistance," *Energy*, vol. 277, Aug. 2023, Art. no. 127675.
[22] A. Gismero, K. Nørregaard, B. Johnsen, L. Stenhøj, D.-I. Stroe, and E. Schaltz, "Electric vehicle battery state of health estimation using incremental capacity analysis," *J. Energy Storage*, vol. 64, Aug. 2023, Art. no. 107110.
[23] R. Xiong, Y. Zhang, J. Wang, H. He, S. Peng, and M. Pecht, "Lithium-ion battery health prognosis based on a real battery management system used in electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4110–4121, May 2019.
[24] X. Feng et al., "Online state-of-health estimation for li-ion battery using partial charging segment based on support vector machine," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8583–8592, Sep. 2019.
[25] B. Gou, Y. Xu, and X. Feng, "An ensemble learning-based data-driven method for online state-of-health estimation of lithium-ion batteries," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 2, pp. 422–436, Jun. 2021.
[26] C. Zhang et al., "Battery SOH estimation method based on gradual decreasing current, double correlation analysis and GRU," *Green Energy Intell. Transp.*, vol. 2, no. 5, Oct. 2023, Art. no. 100108. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2773153723000440
[27] Z. Zhou, Y. Liu, M. You, R. Xiong, and X. Zhou, "Two-stage aging trajectory prediction of LFP lithium-ion battery based on transfer learning with the cycle life prediction," *Green Energy Intell. Transp.*, vol. 1, no. 1, Jun. 2022, Art. no. 100008.
[28] B. Jiang, H. Dai, and X. Wei, "Incremental capacity analysis based adaptive capacity estimation for lithium-ion battery considering charging condition," *Appl. Energy*, vol. 269, Jul. 2020, Art. no. 115074.
[29] E. Kheirkhah-Rad, A. Parvareh, M. Moeini-Aghtaie, and P. Dehghanian, "A data-driven state-of-health estimation model for lithium-ion batteries using referenced-based charging time," *IEEE Trans. Power Del.*, vol. 38, no. 5, pp. 3406–3416, Oct. 2023.
[30] R. Li, J. Hong, H. Zhang, and X. Chen, "Data-driven battery state of health estimation based on interval capacity for real-world electric vehicles," *Energy*, vol. 257, Oct. 2022, Art. no. 124771.
[31] R. Xiong et al., "A data-driven method for extracting aging features to accurately predict the battery health," *Energy Storage Mater.*, vol. 57, pp. 460–470, Mar. 2023.
[32] E. Martinez-Laserna et al., "Battery second life: Hype, hope or reality? A critical review of the state of the art," *Renewable Sustain. Energy Rev.*, vol. 93, pp. 701–718, Oct. 2018.
[33] M. Etxandi-Santolaya, L. Canals, T. Casals Montes, and C. Corchero, "Are electric vehicle batteries being underused? A review of current practices and sources of circularity," *J. Environ. Manage.*, vol. 338, Jul. 2023, Art. no. 117814.
[34] Albatross, "Albatross H2020 EU." Accessed: Feb. 25, 2024. [Online]. Available: https://albatross-h2020.eu
[35] M. Etxandi-Santolaya, L. Canals Casals, and C. Corchero, "Estimation of electric vehicle battery capacity requirements based on synthetic cycles," *Transp. Res. Part D: Transport Environ.*, vol. 114, Jan. 2023, Art. no. 103545.
[36] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
[37] International Energy Agency, "Global EV outlook 2024," Tech. Rep., 2024.

Maite Etxandi-Santolaya received the double master's degree in energy and industrial engineering and the Ph.D. degree in environmental engineering from the Universitat Politècnica de Catalunya, Barcelona, Spain. She is currently a Engineer with the Catalonia Institute for Energy Research in the Energy Systems Analytics group. Her research interests include EV battery state algorithm development.

Tomas Montes is a Industrial Engineer with the University of Oviedo, Oviedo, Spain. During the last three years he has been a Researcher with the Catalonia Institute for Energy Research in the Energy Systems Analytics group where he has specialized in developing battery lifetime models and using them to optimize the use of batteries. In addition, he has experience analyzing the possibilities of the batteries after their first life. In this context, he is carrying out his Ph.D. covering different decision algorithms during the life cycle of the batteries.

Lluc Canals Casals received the Ph.D. degree in modelling li-ion battery aging for second-life business models from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. Then he did a Postdoctorate with the Catalonia Institute for Energy Research until he became a Professor with UPC where he teaches Project Management and continues his research. He defines himself as techno-skeptic. Lluc Canals Casals is a Serra Hunter Fellow.

Cristina Corchero received the Ph.D. degree in statistics and operations research from the Universitat Politècnica de Catalunya, Barcelona, Spain. She is a cofounder of Bamboo Energy and scientific advisor of the Energy Systems Analytics Group with the Catalonia Institute for Energy Research. Her research focuses on applying advanced optimisation tools to energy systems. Sheled and participated in several European, national, regional and industrial projects on energy efficiency and new technologies system integration and has registered and patented several software solutions.

Josh Eichman was an adjunct Professor with the Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA and a Senior Researcher with the National Renewable Energy Laboratory. He is the group lead for Energy Systems Integration with the Catalonia Institute for Energy Research (IREC) in Barcelona, which focuses on better understanding and optimizing the use of current and emerging energy technologies to support the clean energy transition.