Analysis of Li-ion battery degradation using self-organizing maps

Pablo Pastor-Flores
Group of Power Electronics and
Microelectronics (GEPM)
University of Zaragoza
Zaragoza, Spain
pablop@unizar.es

Carlos Bernal-Ruiz
Group of Power Electronics and
Microelectronics (GEPM)
University of Zaragoza
Zaragoza, Spain
cbernal@unizar.es

Antonio Bono-Nuez
Human Openware Research Lab
(HOWLab)
University of Zaragoza
Zaragoza, Spain
antoniob@unizar.es

Iván Sanz-Gorrachategui
Group of Power Electronics and
Microelectronics (GEPM)
University of Zaragoza
Zaragoza, Spain
isgorra@unizar.es

Bonifacio Martín-del-Brío
Human Openware Research Lab
(HOWLab)
University of Zaragoza
Zaragoza, Spain
bmb@unizar.es

Jesús Sergio Artal-Sevil
Gestión estratégica de la energía
eléctrica (ge3)
University of Zaragoza
Zaragoza, Spain
jsartal@unizar.es

Abstract— **This paper proposes a new methodology to identify the different degradation processes of Li-Ion battery cells. The goal of this study is to determine if different degradation factors can be separated by waveform analysis from aged cells with similar remaining capacity. In contrast to other works, the proposed method identifies the past operating conditions in the cell, regardless of the actual State of Health. The methodology is based on a data-driven approach by using a SOM (Self-organizing map), an unsupervised neural network. To verify the hypothesis a SOM has been trained with laboratory data from whole data cycles, to classify cells concerning their degradation path and according to their discharge voltage patterns. Additionally, this new methodology based on the SOM allows discriminating groups of cells with different cycling conditions (based on depth of discharge, ambient temperature and discharge current).**

This research line is very promising for classification of used cells, not only depending on their current static parameters (capacity, impedance), but also the battery use in their past life. This will allow making predictions of the Remaining Useful Life (RUL) of a battery with greater precision.

Keywords— Neural networks, Self-organizing maps (SOM), batteries, battery degradation, Remaining useful life (RUL).

I. INTRODUCTION

Nowadays, the number of devices and systems that require autonomous power (smartphones, sensor networks, electric vehicles, photovoltaic generation systems, off-grid installations) is constantly growing. In this context, it is well-known the importance of the prediction of the state and how batteries age. Batteries degrade due to a large number of factors, such as temperature, charge and discharge currents, range of working voltages, and even the different applications for which they are used. Due to this degradation, the behavior of the same cell varies in time, reducing its capacity, increasing its internal resistance, but also changing its voltage-current behavior in the same application.

In this area, there are different approaches to parameterize the aging of cells. For instance, there is a large scope of study of the chemistries of each cell, in order to understand the degradation processes that take place (i.e [1], [2]).
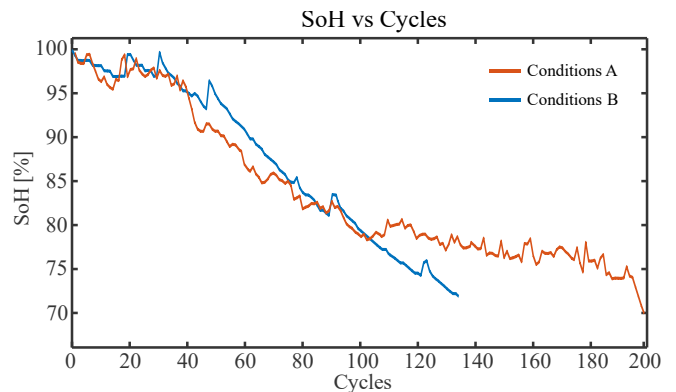

Fig. 1 SoH comparison in battery cycling for two different conditions.

The classical point of view suggests metrics based on characteristics of models for each of the chemical [3]. The most common metric to define the state of a battery is the State of Health (SoH), which yields a percentage based on a quotient for several of the model characteristics (Capacities, impedances and so on). A variety of techniques are listed in [4], [5]. However, this approach to battery aging does not takes into account neither non-linearities nor small differences between equally manufactured cells and the cells' composition. Two cells might have the same SoH at certain moments of their useful life, but being a multidimensional problem, due to their different applications, discharge currents, temperatures or depth of discharge, they can have different aging graphs (Fig. 1).

Another interesting metric is the Remaining Useful Life (RUL). This measure combines SoH with consumption predictions for an application to determine how long a battery can continue to be used in that specific application. The definitions for this metric can be consulted in [6]–[9].

This paper presents the study of battery aging by applying pattern recognition and exploratory analysis techniques to data of several cells with different uses. Our main goal is to classify cells according to the purpose they have served in their past cycles, so that a prediction of the remaining useful life can be given more realistically than with currently used methods. To this end, an unsupervised classifier based on the neural network algorithm Self-organized Maps (SOM) has been used.
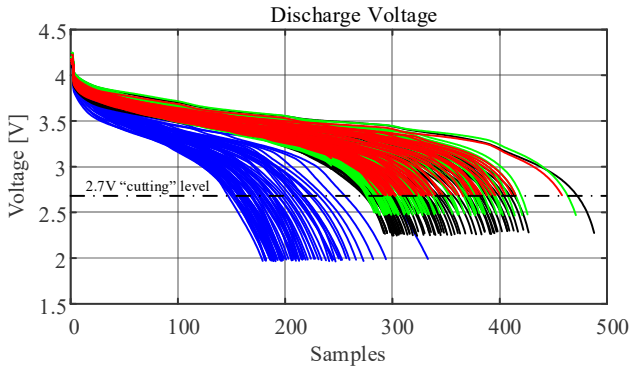
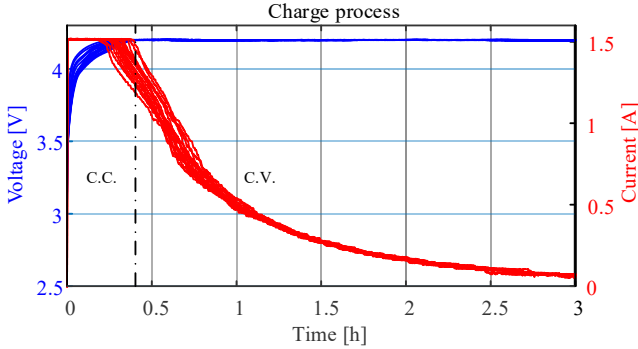Fig. 2 Discharge voltage for 4 different DoD levels

Fig. 3 Voltage and current on charge process

In Section II the database used is presented, the SOM algorithm is explained, and the data groups used on our work are defined. Results are shown and discussed in Section III. Section IV presents the conclusions and establishes future work to be conducted.

II. METHODOLOGY

A. Database

A group of 34 Li-ion 2 Ah 18650 batteries have been selected. The batteries were cycled by NASA Ames Prognostics Center of Excellence (PCoE) ([7], [10]). They were cycled until a reduction of 20-30 % on the End of Life (EoL) was reached (criterion commonly used in these studies). Charge profiles are similar for all cells, with a constant current (CC) phase at 1.5 A until reaching 4.2 V, and a constant voltage (CV) phase until current drops to 20 mA, Fig. 3. Discharge profiles were made for different values on discharge current (1 A, 2 A y 4 A), temperature on the climatic chamber (4 ºC, 24 ºC y 43 ºC.) and depth of discharge (2 V, 2.2 V, 2.5 V y 2.7 V).

Discharge voltage waveforms have been selected for this study (Fig. 2) since charge patterns were similar for all the cells. Current waveforms in the discharge were controlled (CC phase), and only showed variability according to the battery health in the constant voltage zone (CV) (Fig. 3). As cells of this study have different DoD, and introducing this information to the neural network could make the clusters depend on this data (which is not the object of our study), it was decided to "cut" all the voltage curves analyzed at 2.7 V, the shallower DoD level (Fig. 2).
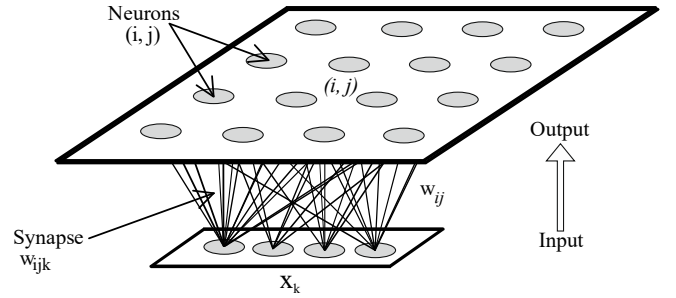

Fig. 4 Generic SOM architecture scheme. A 4x4 map is represented.

SOM also requires constant length vectors; however, discharge waveforms have a different duration depending on the discharge current and its state of degradation. For this reason, it was decided to make discharge curves independent of time and capacity by normalizing them according to the state of charge (SoC), defined as

$$SoC(t) = 100 - \frac{Q_{DISCHARGED}(t)}{Q_{TOTAL}} \cdot 100 \qquad (1)$$

Where $Q_{DISCHARGED}(t)$ represents the charge extracted from the battery at time t of the cycle, calculated by Coulomb-Counting [11] and Q_{TOTAL} represents $Q_{DISCHARGED}(t)$ evaluated at the end of the cycle (2.7 V).

Thus, the SOM does not receive any information about the time duration of the discharge or the capacity extracted. This normalization has been carried out by linearly interpolating the discharge voltage curves to obtain a vector of one hundred points corresponding to one hundred percentage points of the SoC, as indicated by equation (2)

$$V_x = V_a + (Q_x - Q_a) \cdot \frac{V_b - V_a}{Q_b - Q_a} \qquad (2)$$

Where V_x represents interpolated voltage; Q_x is the capacity of the point to be interpolated, calculated by extracting a hundred equidistant points in the evolution of the cell capacity in the cycle; and V_a, V_b, Q_a, Q_b are voltages and capacities of the upper and lower points to the point to interpolate.

Finally, incoherent information and incomplete discharge patterns have been deleted from the database. Two kinds of spurious data have been found, discharge curves not reaching the DoD established (experiments stopped at a higher DoD than expected) and curves of extremely short duration (due to unknown reasons, experiments were stopped too early). This type of input data (spurious, outliers) is usually eliminated in data analysis techniques.

B. Self-organizing maps (SOM)

The SOM neural model (Self-organizing Maps) [12] (Fig. 4) has been selected for the analysis. This algorithm is more useful for this application than common clustering techniques, such as k-means, because the SOM projects the high-dimensionality input space to a map, which allows visualizing the trajectory drawn by a succession of temporal patterns associated with the same data source (in this case to the same cell). Thus, the SOM enables us to observe the trajectory that

a cell draws on the map surface through its voltage patterns, showing the degradation process.

The SOM is a competitive network, with an input layer and an output layer (the map) of unsupervised neurons, used for pattern recognition, cluster search and database visualization. The SOM has two modes of operation: training and inference. Given a map of $nx \times ny$ neurons with an input layer of n variables x_k $(1 \leqslant k \leqslant n)$, each neuron (i, j) stores a vector of synaptic weights w_{ij} of n components (Fig. 4). In inference, each neuron (i, j) calculates the similarity between the input vector x_k and its weights w_{ijk}. The Euclidean distance is commonly used:

$$d^2(w_{ij}, x) = \sum_{k=1}^{n} (w_{ijk} - x_k)^2 \qquad (3)$$

The neuron whose weights minimizes the distance with the input vector is considered to have recognized the input pattern and is called "winner" or "best-matching unit" (BMU).

In the training phase the synaptic weights are adjusted. Starting with a set of initial (random) weights, in each iteration t an input vector $x(t)$ is presented and the search process of the BMU is performed as in the inference phase. The weights of the BMU and those around it (neighborhood) are updated:

$$\triangle w_{ijk}(t) = \varepsilon(t) \cdot (x_k(t) - w_{ijk}(t)) \qquad (4)$$

Where $\varepsilon(t)$ corresponds to the learning rate, which decreases with time:

$$\varepsilon(t) = \varepsilon_0 + (\varepsilon_f - \varepsilon_0)\frac{t}{t_f} \qquad (5)$$

ε_0 represents the initial learning rate ($\varepsilon_0 < 1.0$), ε_f is the final one ($\varepsilon_f \approx 0.01$) and t_f is the number of iterations to reach ε_f.

In applications with very different input variables with very different ranges, the input data should be normalized (for instance, between [0,1] or [-1,1]). However, for this application it has been decided not to normalize the data because all the inputs are of the same type (voltages) with similar ranges (between 4.1 V and 2.7 V).

This training algorithm generates a non-linear projection of the multidimensional input space (database) onto a two-dimensional map, allowing to visualize the structure of the database and the possible clusters presented in it. More information about this algorithm can be found in [13]–[15]. In this work, the SOM Toolbox developed for MATLAB® by the Kohonen research group has been used [16].

C. Selection of study groups

In order to define study groups from the database, with variations in controlled cycling conditions, several cell-groups with similar experimental conditions have been selected from the NASA database, with the aim of obtaining clearer conclusions.

The NASA database is organized in groups of three or four cells with constant ambient temperature and discharge conditions for the group, but a different DoD for each cell. The database presents nine groups of cells with different parameters, however, after eliminating those groups with large amounts of spurious, non-constant discharge currents, or groups of cells with more than one varying characteristic,

TABLE 1
Studied cells parameters

Test	Cells	V min [V]	Temp. [ºC]	I disch. [A]	N. disch. Cycles
1	22 units*	2 | 2.2 | 2.5 | 2.7	4 | 24 | 43 | 24/44	1 | 2 | 4	40-170
2	5 | 6 | 7	2.7 | 2.5 | 2.2	24	2	170
3	45 | 46 | 47 | 48	2 | 2.2 | 2.5 | 2.7	4	1	72
4	29 | 30 | 31 | 32	2 | 2.2 | 2.5 | 2.7	43	4	40

* Test 1: all the cells excluding spurious cells and cells with non-constant discharge current

only three groups are considered useful for our study. Each cell has between 40 and 170 discharge cycles. Additionally, a test is carried out with all the cells to observe the map organization according to the characteristics of DoD, ambient temperature and discharge current. These four groups are in TABLE 1. The original names of the cells have been maintained to facilitate the reproducibility of our experiments.

Experiments will be carried out for each group defined, introducing to the SOM the discharge voltage curves, obtained as detailed in Section II.A.

After some preliminary simulations, an 18 × 10 SOM is considered adequate to clearly see and separate the cell clusters. The weight vectors w_{ij} have been initialized with random values obtained from a fixed seed, in order to perform several tests with the same initial conditions.

In the resulting maps showed in this paper (e.g. Fig. 5), the so-called U-matrix visualization has been added (first map in every figure), which represents the distance of the vector weight of each of the map's neurons with their neighbors. Neurons in blue represent similar neurons (with similar weights), while yellow neurons represent greater inter-neuron distances, suggesting borders between natural clusters.

III. RESULTS AND DISCUSSION

The input vector of each training example (each discharge cycle) is composed of 100 voltage points corresponding to the percentual SoC points. In inference, the SOM projects the input vectors onto the map surface (neurons); in (Fig. 5) it is represented the U-matrix and three maps with the cell projections. These three maps are extracted from the same trained SOM but representing onto every neuron the projection of every input vector, coloring its ambient temperature, depth of discharge and discharge current, respectively (Fig. 5).

Our first test consists of training a SOM with all the cells available in the NASA database. In this case, we want to validate the SOM ability to differentiate cells with different usage characteristics recognized as crucial in the study of aging. The resulting clusters (Fig. 5) have clear and polarized results for the discharge current and ambient temperature maps. For instance, in the case of the current discharge map, a gradient is observed from the upper left to the lower right (decreasing current).
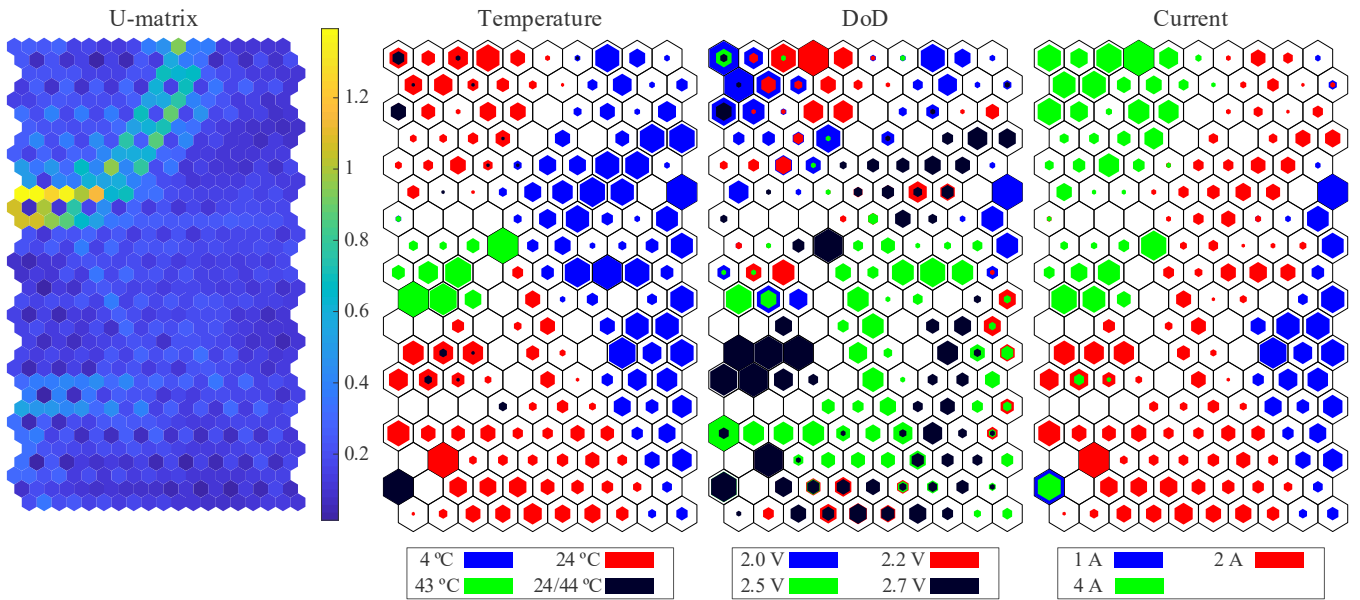
Fig. 5 SOM trained with all the database cells (Test 1). The first map is the U-matrix; the other three maps are the same trained SOM but representing the projection onto the map of the different cells (input vectors), colored depending on their temperatura, DoD and discharge current, respectively.
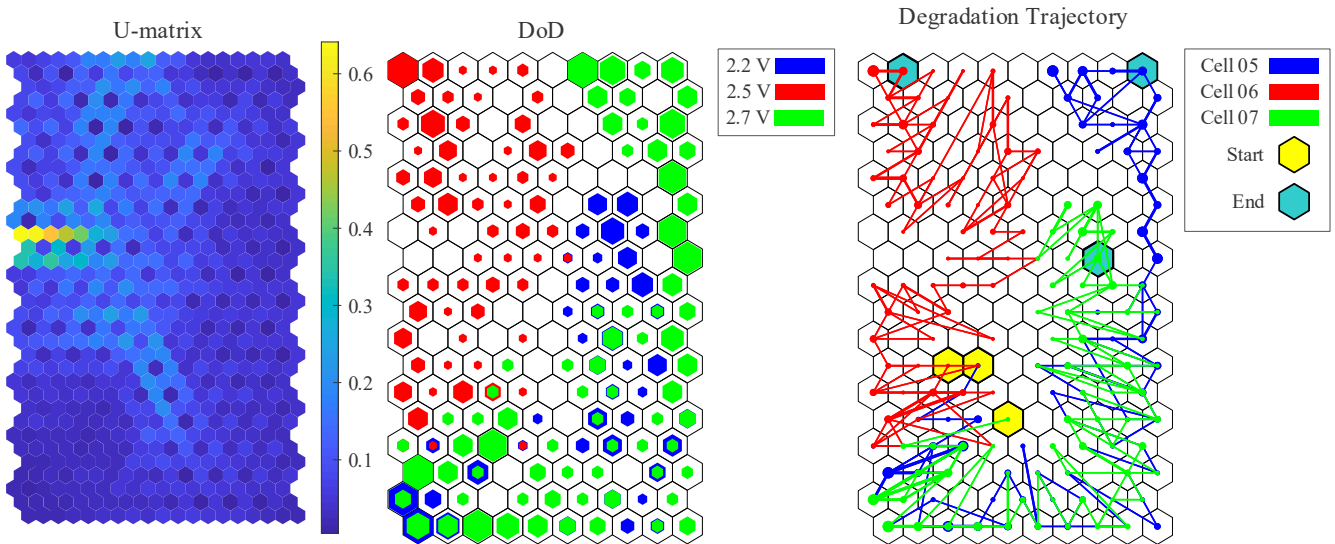

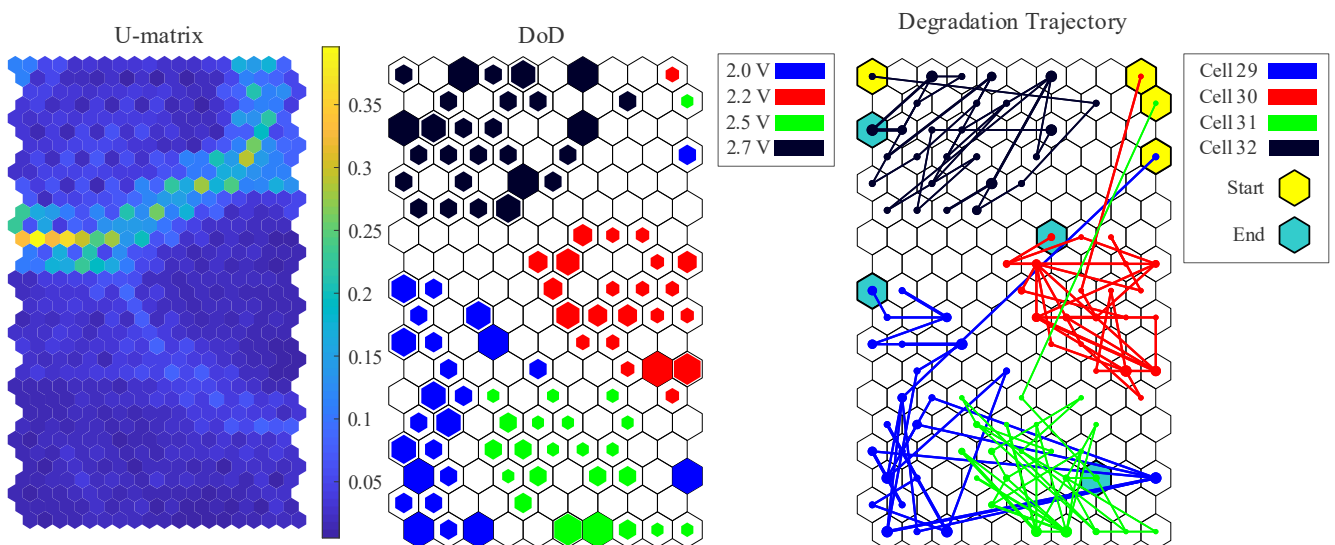Fig. 6 Test 2: Three cells with 170 cycles, Ambient Temperature of 24ºC and Current of 2 A.

Fig. 7 Test 4: Four cells with 40 cycles, Ambient Temperature of 43 ºC and Discharge Current of 4 A.
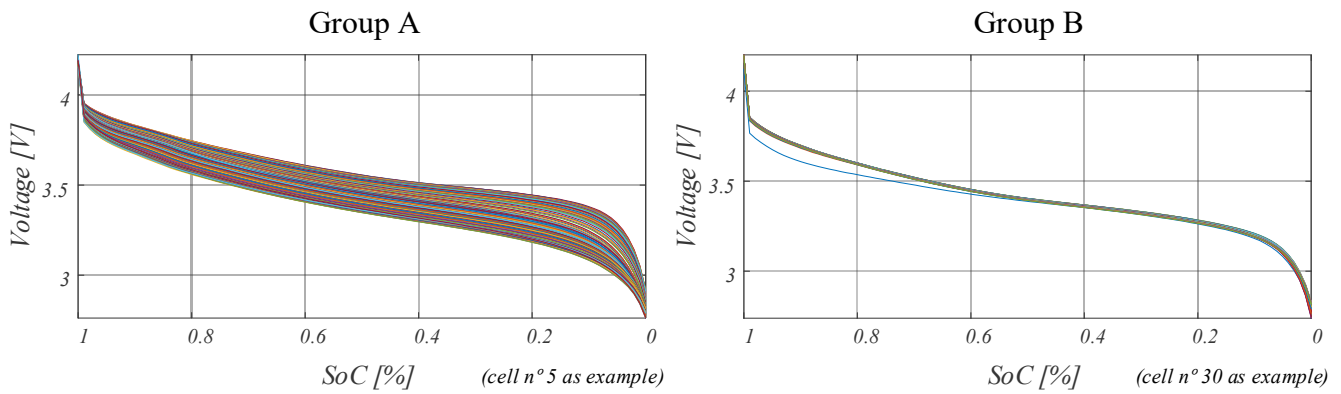
Group A

Group B

Fig. 8 Examples of analyzed discharge voltage curves. Group A represents cells with normal degradation meanwhile cells with premature degradation have curves like Group B.

In the case of temperature map, clear clusters appear for 4 ºC and 43 ºC; however, for the temperature of 24 ºC, two different clusters appear. The upper left corresponds to the group for the cells discharged at 4 A. In the DoD map, groupings coincidences are observed mostly between the cells discharged at 2.5 V and 2.7 V in the lower zone, and 2.0 V and 2.2 V in the upper area of the map. However, these groups seem to depend on their currents and temperatures, as can be shown on current and temperature maps. A clear border appears in the U-matrix in the upper left zone, coinciding with the cells discharged at 4 A. Likewise, a subgroup of cells discharged at 4 A and 43 ºC is present, located in the left-middle zone of the map, representing the most aggressive discharges with degradation effects. The cells discharged at variable temperatures of 24 ºC and 44 ºC between cycles are scattered in the temperature zones of 24 ºC.

Tests 2, 3 and 4 use cells with the same current and temperature conditions, but different DoD levels. In the voltage map of test 4, clear clusters are differentiated for each DoD level. (Fig. 7). However, at test 2, cells with DoD of 2.2 V and 2.7 V are not clearly discriminated. The results of test 3 are very similar to test 2, so detailed graphs have not been included.

The trajectory over the map of each of the cells involved in the experiment has been included in these tests, painting the winning neurons for each of the voltage patterns introduced. The start and endpoints have been labeled to define differences between Beginning of Life (BoL) and End of Life (EoL) points.

Start points (yellow color) of each of the trajectories are located in contiguous cells in all the tests, a symptom that their behaviors in the first life cycles were very similar. As its useful life is consumed, it is observed how the trajectories are differentiated on the map (until its final point, in blue) depending on the specific use of each cell.

In Fig. 6 a very clear life trajectory for 2.5 V DoD cell can be distinguished, however, the behavior of the cells discharged up to 2.2 V and 2.7 V seems very similar until approximately 70 % of life useful consumed of both cells (using data labels to identify this point).

In test 4, Fig. 7, with cells discharged at 4 A and 43°C, different clusters can be observed for DoD levels, although in this case, the trajectory of each of the cells does not provide temporal information that can help predict the remaining useful life.

Studying the voltage curves introduced to the SOM for each of the experiments, it can be seen that in the latter case, trajectories have similar behavior to those shown in Group B of Fig. 8. This group, cycled with such extreme temperature and current condition (43 ºC and 4 A), defined EoL is reached with only forty charge-discharge cycles. In addition, it can be observed in Fig. 8, that, for the case represented in Group B, behavior changes in voltage curves between cycles are not observed.

These results could appear due to premature cell death because of high temperature and current conditions. The rest of the studied cells had similar behaviors to Group A in their evolutions in the discharge voltage curve.

IV. CONCLUSIONS AND FUTURE WORK

In this work, a new technique for determining different usage patterns associated with battery aging has been proposed. A machine learning algorithm, SOM, has been used to find clusters of similar cell patterns and visualize them onto a map. Also, the SOM allows seeing the temporal evolution of a battery cell, as trajectories onto the map surface. Tests have been carried out to approximate the remaining useful life of the studied cells by using the projected trajectories by the discharge voltage curves in a battery cycling database.

This work is an initial study and the results provided are qualitative, but after analyzing our results, the ability of the SOM to discriminate ambient temperature conditions and discharge current in groups of cells with very heterogeneous conditions has been verified. The SOM is also capable of presenting differentiations between DoD levels for groups of cells with similar temperature conditions and discharge currents.

In addition, the trained SOM algorithm is able to determine acceptably distinguishable trajectories to estimate the remaining useful life under similar cycling conditions, for groups of cells with a large number of cycles (greater than one hundred discharges) and with some variability in their discharge voltage waveforms due to the effect of degradation.

It should be noted that, in spite of having a reduced study dataset and a limited number of voltage curves for each of the samples (22 cells with 40-170 discharge cycles per cell), it has been possible to verify the possibility of obtaining distinguishable trajectories with degradation information in the use for different applications, by using as input SOM inputs only the information related to the discharge voltage.

Although these are preliminary results, future applications are very promising. With the increase of autonomous power applications, the need to determine the real state of health of the batteries will be a crucial aspect in the sizing of facilities and the health diagnosis for the reuse of batteries.

The distinction of cells used in different applications by using only their discharge voltage patterns would allow classifying batteries depending on the usage they were intended to, and not only by using static metrics, such as SoH, which does not provide complete information about their health state.

Likewise, being able to locate the state of life of a cell within a specific trajectory, would allow not only to establish its past life, but also to estimate its remaining useful life. In addition, it should be noted that for the realization of these tests is only required to perform a simple cell cycle, which does not require complex equipment.

Currently, we are looking for a more complete database with more cells and more conditions, which allows to confirm the results found in this work and to deepen the new proposed method.

REFERENCES

[1] B. Sood, M. Osterman, and M. Pecht, "Health monitoring of lithium-ion batteries," in *2013 IEEE Symposium on Product Compliance Engineering (ISPCE)*, 2013, pp. 1–6.

[2] L. S. Kanevskii and V. S. Dubasova, "Degradation of lithium-ion batteries and how to fight it: A review," *Russ. J. Electrochem.*, vol. 41, no. 1, pp. 1–16, Jan. 2005.

[3] J. Zhang and J. Lee, "A review on prognostics and health monitoring of Li-ion battery," *J. Power Sources*, vol. 196, no. 15, pp. 6007–6014, Aug. 2011.

[4] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 572–587, Apr. 2016.

[5] B. Balagopal and M. Y. Chow, "The state of the art approaches to estimate the state of health (SOH) and state of function (SOF) of lithium Ion batteries," in *Proceeding - 2015 IEEE International Conference on Industrial Informatics, INDIN 2015*, 2015, pp. 1302–1307.

[6] L. Liao and F. Kottig, "Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction," *IEEE Trans. Reliab.*, vol. 63, no. 1, pp. 191–207, Mar. 2014.

[7] B. Saha, K. Goebel, and J. Christophersen, "Comparison of prognostic algorithms for estimating remaining useful life of batteries," *Trans. Inst. Meas. Control*, vol. 31, no. 3–4, pp. 293–308, Jun. 2009.

[8] M. Pajovic, P. Orlik, and T. Wada, "Remaining Useful Life Estimation of Batteries using Dirichlet Process with Variational Bayes Inference," in *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, 2018, pp. 2729–2735.

[9] A. O. D. Approach, J. Zhou, D. Liu, Y. Peng, and X. Peng, "Dynamic Battery Remaining Useful Life Estimation : An On-line Data-driven Approach," pp. 0–3, 2012.

[10] K. Goebel and B. Saha, "DASHlink - Li-ion Battery Aging Datasets," 2010. [Online]. Available: https://c3.nasa.gov/dashlink/resources/133/. [Accessed: 04-Mar-2019].

[11] K. S. Ng, C. S. Moo, Y. P. Chen, and Y. C. Hsieh, "Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries," *Appl. Energy*, vol. 86, no. 9, pp. 1506–1511, Sep. 2009.

[12] T. Kohonen, *Self-Organizing Maps*, 3rd ed. Springer-Verlag Berlin Heidelberg, 2001.

[13] B. Martín-del-Brío, A. Bono-Nuez, and N. Medrano-Marqués, "Self-organizing maps for embedded processor selection," *Microprocess. Microsyst.*, vol. 29, no. 7, pp. 307–315, Sep. 2005.

[14] O. Simula and E. Alhoniemi, "SOM based analysis of pulping process data," Springer, Berlin, Heidelberg , 1999, pp. 567–577.

[15] T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas, "Engineering applications of the self-organizing map," *Proc. IEEE*, vol. 84, no. 10, pp. 1358–1384, 1996.

[16] Helsinki University of Technology, "SOM Toolbox Software Library 2.0 for MATLAB." [Online]. Available: http://www.cis.hut.fi/projects/somtoolbox/. [Accessed: 05-Mar-2019].