Hybrid Modelling for Second-Life Battery Optimization in Light Electric Vehicles

Andrés Bernabeu-Santisteban[1*], Alejandro Clemente[1,2], Bernhard C. Geiger[3,4], Franz M. Rohrhofer[3], Francisco Díaz-González[2] and Lluís Trilla[1]

1 Power Systems, Catalonia Institute for Energy Research, Sant Adrià de Besòs, 08930 Barcelona, Spain
2 Polytechnic University of Catalonia, Jordi Girona 31, 08034 Barcelona, Spain
3 Know Center Research GmbH, Sandgasse 34, 8010 Graz, Austria
4 Signal Processing and Speech Communication Laboratory, Graz University of Technology, Inffeldgasse 16c, 8010 Graz, Austria
**Corresponding author. abernabeu@irec.cat*

ABSTRACT

The repurposing of second-life lithium-ion batteries (LIBs) offers a sustainable solution for less demanding applications, such as light electric vehicles (LEVs) or stationary storage. However, limited information on their usage history and internal condition hinders their effective reuse. This paper presents a hybrid modelling (HM) approach that combines a data-driven model (DDM) with a physics-based model (PBM) to assess battery health and predict future performance. The DDM estimates, from a simple charge test, three key parameters: state of health, full equivalent cycles, and active material volume fractions. These are then used to initialize the PBM, which simulates degradation and estimates remaining useful life under LEV application profiles. Trained on experimental data from NMC811 cells, the HM demonstrates strong potential as a practical tool for second-life battery evaluation for different types of cells, integrating adaptability and physical accuracy in a single framework.

https://doi.org/10.25364/978-3-903374-48-5_26

1. INTRODUCTION

The growing integration of renewable energy sources and the electrification of transport have significantly increased the demand for reliable and efficient energy storage systems (ESS). Among the various available technologies, lithium-ion batteries (LIBs) have emerged as the dominant solution due to their high energy density, long cycle life, and lower production costs (Nitta et al., 2015). One of the main advantages of LIBs is their versatility across a wide range of applications. They can be used in both stationary systems, such as grid-scale energy storage, and mobile applications, including consumer electronics and electric vehicles (EV).

In demanding applications such as EVs, LIB packs are generally considered unsuitable when their available capacity falls below 80% of the nominal. This threshold typically marks the end of first-use life, although most cells within the pack still retain sufficient capacity and functionality for less demanding applications such as light electric vehicles (LEVs) or off-grid ESS. Reusing these cells in second-life batteries (SLBs) offers significant environmental and economic benefits by reducing resource consumption, even though it is a time-consuming and complex process.

Despite the promising idea of giving used batteries a second life, the process is not yet optimized, and the deployment SLBs is hindered by several challenges. First, a major barrier is the lack of data provided by the manufacturers, such as cell characteristics and pack specifications. Additionally, first-life usage information, like operating conditions and degradation patterns, are often unavailable and critical for selecting an appropriate second-life usage. This lack of transparency complicates accurate assessment of battery state and forecasting, increasing uncertainty and risk (Von Bülow et al., 2024).

To bridge this gap, battery modelling and data analysis have become essential tools for inferring the internal state of LIBs from measurable data, enabling both diagnostics and prognostics. Physics-based models (PBMs) and data-driven models (DDMs) have gained significant attention and have emerged as powerful approaches for evaluating and forecasting the behaviour of SLBs.

PBMs are grounded in electrochemical and physical laws, with various model typologies depending on the underlying assumptions. However, these models are typically computationally intensive, require extensive parameterization, and lack adaptability when applied to batteries already degraded or with unknown composition (Jokar et al., 2016). In contrast, DDMs rely on statistical or machine learning techniques to learn patterns from measured data. They

can model complex energy system dynamics; however, they often struggle to generalize beyond their training data.

To leverage the strengths of both approaches, hybrid models (HMs) and physics-informed machine learning (PIML) have emerged as a promising strategy. These models integrate physical laws and machine learning in a unified framework, enabling both accurate internal state estimation and generalizable prediction of battery performance. By embedding physics-based constraints or using machine learning to estimate hard-to-measure physical parameters, PIML provide a balanced trade-off between accuracy, generalizability, and interpretability (Navidi et al., 2024).

Recent advances in hybrid modelling include the use of DDMs to estimate parameters for PBMs and using these last ones to generate synthetic data (Wang et al., 2024). However, a major research gap still exists in the application of HM approaches for SLBs, particularly in cases where no prior usage history is available. Most existing studies either focus exclusively on new cells under controlled conditions or rely on assumptions that are not valid for cells with unknown degradation pathways.

In this context, the present work proposes a HM, comprised of a DDM and a PBM, designed to address the uncertainties associated with SLBs. Firstly, the DDM based on XGBoost, estimates health indicators, such as state of health (SOH), full equivalent cycles (FEC), and active material volume fractions ($\varepsilon_{act,k}$), from a single laboratory charge test. Then a PBM, based on the pseudo-two-dimensional model (P2D) and incorporating degradation mechanisms, simulates future battery performance, predicting SOH evolution and remaining useful life (RUL) based on DDM outputs. By integrating both modelling approaches, the HM enables informed decision-making for SLBs deployment, even in the absence of historical data. The model provides valuable insights into the battery's current condition and estimates its suitability for specific applications such as LEVs.

The rest of this paper is structured as follows: Section 2 details the development and integration of the DDM and PBM components. Section 3 describes the experimental calibration and validation of the models using NMC811 cells. Section 4 outlines the methodology for performance estimation under realistic use cases. Section 5 presents the results, and Section 6 concludes with insights and future directions for SLBs modelling.

2. HYBRID INFORMED MODEL FORMULATION

As previously noted, limited information on LIB characteristics and usage history during their first life represents a major barrier to the optimal deployment of SLBs. The absence of such data and knowledge about the battery's state increase the risks associated with operation and pose significant challenges in determining its most suitable application. To address this issue, the article proposes a HM capable of providing a clearer assessment of actual LIB condition and estimating the future performance under a certain application, with just a full charge test.

The designed HM consist of two complementary components: a DDM responsible of assessing the current battery condition and a PBM that estimates future performance under second-life applications. To achieve that objective, the DDM provides three key indicators: SOH, which quantifies the loss of battery capacity relative to the nominal value; the FEC, which reflects battery usage expressed on cycles; and the $\varepsilon_{act,k}$, a parameter representing the amount of active material available in the electrodes per unit volume. This latter parameter determines cell capacity in PBMs and provides valuable insights into the internal states of the LIB (Chen et al., 2020).

Based on the current state analysis provided by the DDM, the PBM estimates battery performance in second-life applications. Specifically, the model predicts SOH evolution and generates additional diagnostic indicators such as the RUL, expressed in years, which quantifies the time left before the battery reaches its end of life or becomes unsuitable for the studied second-life application (O'Kane et al., 2022).

HM is designed to operate with cell-level data, focusing exclusively on NMC811 cells, one of the most widely used LIB chemistries currently available (Huotari et al., 2021). The model does not, however, assess risk or account for mechanical deformation from stresses or vibrations during the first life, which may cause discontinuous operation and deviations in results.

2.1 DATA DRIVEN MODEL (DDM)

DDMs are widely used across different types of applications and there are multiple options in terms of model types, such as random forest, neural networks (NN), etc. The selection of the model type depends on the intrinsic characteristics of the system aimed to be represented and the characteristics

of available data. For example, time series data are often modelled using recurrent NNs, while tree-based models perform superior for tabular data.

In this study, the objective is to build a DDM capable of estimating SOH, FEC, and $\varepsilon_{act,k}$, from a charge test executed at the laboratory, going from the minimum to the maximum voltage specified by the manufacturer.

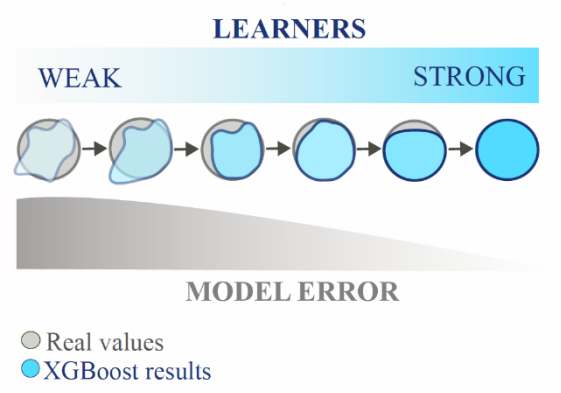

Figure 1: Fundamental operation of a XGBoost algorithm.

For this study scenario, the XGBoost algorithm has been selected, because it works well with tabular data and has been successfully used for predicting battery behaviour with time-dependent variables (Huotari et al., 2021). Moreover, since not all inputs of the developed DDM will be normalized, XGBoost offers a particularly interesting choice.

Like other tree-based models, XGBoost generates outputs from the combined response of multiple individual models but enhances this approach with a boosting method that executes models sequentially. As shown in Figure 1, the initial models, or weak learners, have higher error, while subsequent models learn from previous ones, gradually improving accuracy until forming a precise strong learner. Detailed functions and parameters governing its operation are omitted here but can be found at O'Kane et al. (2022).

2.2 PHYSICS-BASED MODEL (PBM)

PBMs are based on physics and chemical principles, including Fick's laws of diffusion, electrochemical kinetics like the Butler-Volmer equation, and electrical transport laws such as Ohm's law. This foundation makes them robust and typically involve solving multiple partial differential equations (PDEs).

Their main advantage lies in accurately representing the internal states of LIBs, such as current distribution and lithium concentration within electrodes and electrolyte.

Among the existing PBMs, the P2D is one of the most used. It considers two dimensions: the particle radius (r) and the overall location on the battery dimension (x), as shown in Figure 2.

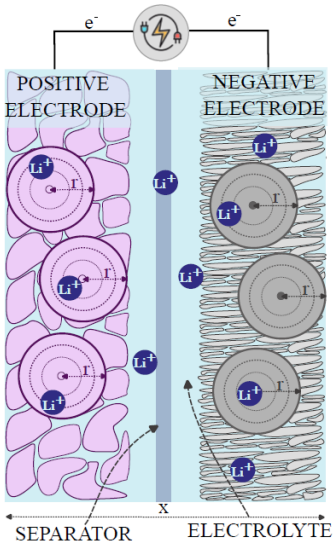

Figure 2: LIB from the perspective of a P2D model

It uses five PDEs and requires 35 parameters to accurately represent the cell performance. Further information of the fundamentals and mathematical background of this model can be found at Chen et al. (2020).

As mentioned before, $\varepsilon_{act,k}$, $k \in \{n, p\}$ is one of the parameters that determines the cell capacity in the P2D. Positive and negative electrodes have different active material volume fractions, denoted as $\varepsilon_{act,p}$ and $\varepsilon_{act,n}$, respectively.

Degradation mechanisms, associated with capacity fade, can also be incorporated into PBMs. In this study, the equations proposed by O'Kane et al. (2022) have been implemented, which model phenomena such as solid electrolyte interface growth, lithium plating, particle cracking and loss of active material.

2.3 INTERACTION BETWEEN MODELS

After outlining in a general manner the models comprising the solution, the operation of the HM is described at Figure 3. The second-life cell is subjected to a charge profile test, during which it is charged from its minimum to maximum voltage, thereby going from 0% to 100% state of charge (SOC). This data generated at the laboratory, is then processed by the DDM which estimates the SOH, FEC, $\varepsilon_{act,p}$ and $\varepsilon_{act,n}$. These last two parameters, that represent the available active lithium within the positive and negative electrodes, are then introduced to the PBM. With this information, the PBM estimates the cell degradation under a specific future usage pattern anticipated in the second-life application.

By repeating the usage pattern several times through simulation, the PBM estimates the expected operational lifespan, by providing the SOH evolution and RUL. In this way the HM enables an informed assessment of the suitability of the cell for a continued use in a new application, reducing the risk and the resources required.

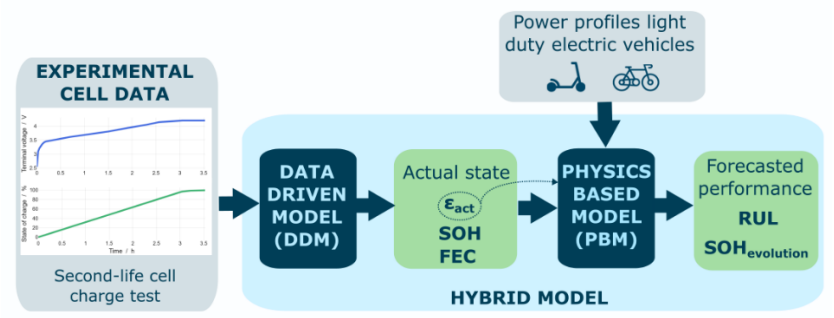

Figure 3: HM structure and principal interactions between DDM and PBM.

3. CALIBRATION AND TRAINING

To properly calibrate the models, experimental tests have been carried out with cylindrical Li-ion cells. The cells modelled are INR21700 M50 (LG-M50), with their main specifications listed in Table 1.

Table 1: INR21700 M50 (LG-M50) principal manufacturer specifications

Item	Manufacturer specifications
Chemistry	NMC811
Nominal capacity	5 Ah
Geometry	Cylindrical
Voltage limits	2.5-4.2V

In total six cells have been considered with different usage patterns: two of them were subjected to driving conditions and the remaining four were charged and discharged according to manufacturer maximum constant current rates, two between 0 to 100% SOC and two between 20 to 80% SOC, respectively. Temperature was held constant at 35ºC.

The six LG-M50 underwent a Reference Performance Test (RPT) at 25ºC after every 25 cycles. This RPT consisted of fully charging and discharging the battery two times at a c-rate of C/3, to track the degradation in the same conditions.

3.1 DDM

The DDM was trained on charge profiles obtained during the RPT using the XGBoost package (Chen & Guestrin, 2016) implemented in Python. The objective is to enable the assessment of the SOH for any NMC811 cell, regardless of its characteristics, geometry or nominal capacity. To ensure that the DDM is compatible across different cell formats and types, all input features provided to the XGBoost algorithm have been normalized. It is assumed that manufacturing differences between cells do not substantially influence battery performance, hence the model inputs are chosen as listed in Table 2.

Table 2: Inputs necessary for the DDM

Variable	Description
Terminal Voltage (V)	Measured voltage
Temperature (ºC)	Measured cell temperature
C-rate (A/Ah)	Measured current divided by the nominal capacity of the cell
dQdV (Ah/V)	Differential capacity with respect differential voltage
dVdt (V/s)	Differential voltage with respect differential time
Cnormalised (Ah/Ah)	Tracked capacity divided by the nominal capacity of the cell

For the training process SOH, FEC, and $\varepsilon_{act,k}$, have also been provided to the DDM. The experimental SOH was calculated after each RPT, by dividing the maximum capacity at 100% SOC by the nominal capacity specified by the manufacturer. The FEC were computed as the sum of the total discharged and charged energy, divided by twice the nominal capacity of the battery.

The $\varepsilon_{act,k}$ were correlated with the SOH based on PBM simulations. Chen et al. (2020) determined initial values for a new LG-M50 cell, as $\varepsilon_{act,p} = 66.5\%$ and $\varepsilon_{act,n} = 75\%$, respectively. By reducing both active $\varepsilon_{act,p}$ and $\varepsilon_{act,n}$ according to the SOH, a relation between active material volume fractions and capacity decay was found as shown in ***Figure 4.***

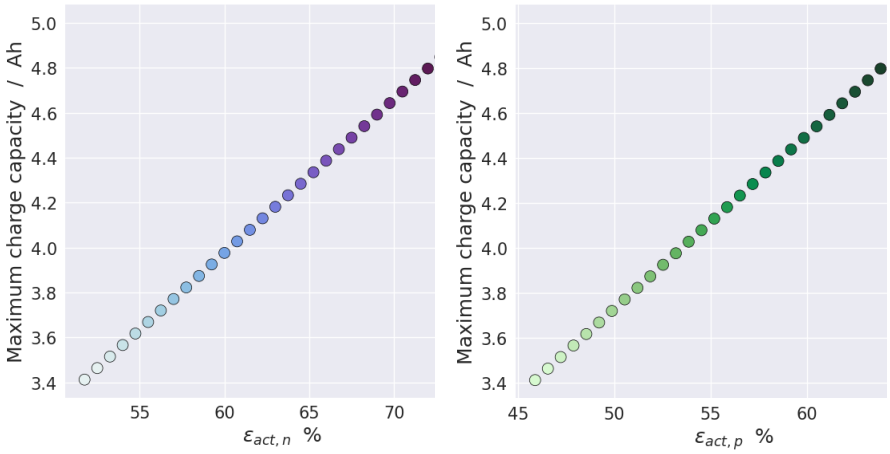

Figure 4: Relation between $\varepsilon_{act,k}$, SOH and maximum charge capacity, according to PBM results. **Left)** Negative electrode $(\varepsilon_{act,n})$ **Right)** Positive electrode $(\varepsilon_{act,p})$.

Using these inputs and targets, the XGBoost model was trained to minimize the squared error between each true target values and the model estimate. Parameters such as the sub-sampling rate (0.8), the fraction of features to evaluate (0.8), and the random seed (42) have been taken from Huotari et al. (2021). A detailed description of the XGBoost model parameters and configuration will be provided in a forthcoming publication and is not included here due to space limitations.

3.2 PBM

The P2D relies on 35 parameters to properly represent the behaviour of LIBs. Therefore, an exhaustive study of the different components comprising the battery is necessary. The values provided by Chen et al. (2020), validated in multiple studies and adequate for the NMC811 chemistry, have been considered in this work. In addition, the degradation equations considered in the P2D, developed by O'Kane et al. (2022), also apply to the same cell. The PBM was developed with the Python library PyBaMM (Sulzer et al., 2021).

The LIBs tested at the laboratory are the same as the ones investigated by both studies, with the same chemistry NMC811 and components. The only parameter modified in this case study was the active material volume fraction,

provided by the DDM, effectively adjusting the battery's maximum capacity according to the state of degradation.

4. METHODOLOGY

After calibrating and training both models, it is possible to test the whole operation of the HM. For validation purposes, a LIB has been tested in the laboratory, for which cycle life, SOH and FEC are known and experimentally determined. This cell, functioning as SLB, is of the same type as those employed during the training phase (LG-M50), however its cycle life data was excluded from training to prevent overfitting the DDM. The cell exhibits a degradation pattern distinct from the training cells, enabling a more realistic evaluation.

To assess the impact of using the SLB in a new application, an electric scooter usage profile was analysed. The LEV was monitored over 24 h in a commuting scenario in the city of Barcelona (Spain), recording battery voltage, current, SOC, and temperature. ***Figure 5*** presents the voltage of one cell in the pack. After executing the DDM, this profile was applied to the PBM and repeated multiple times to estimate future battery's performance under specified SLB conditions.

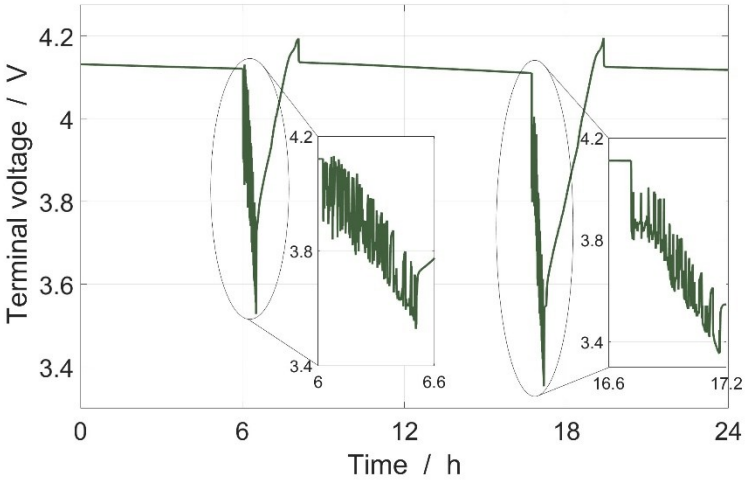

Figure 5: Cell voltage during 24 h of electric scooter operation. The LEV was charged to 100% SOC immediately after use and remained at rest when not in operation.

5. RESULTS

In the first phase, the HM processes the variables listed in Table 2, derived from a charge profile, and uses the DDM to assess the current battery state. Since the data was obtained from a laboratory-tested battery, the SOH and FEC were available for different degradation levels, allowing direct comparison between measured performance and model predictions.

Figure 6 a) and ***b)*** compare the DDM estimation of SOH and FEC values, with the relative error against the measured data. For both variables, the XGBoost model shows lower accuracy when the battery is at the beginning of its cycle life (94–92% SOH) and improved precision after approximately 200 cycles. SOH estimation remains accurate, as it is directly related with battery capacity, with relative errors between 1.1% and 0.3%. In contrast, FEC predictions deviate substantially from actual values, particularly at low FEC. Indeed, this parameter is harder to forecast because it is influenced by multiple variables such as usage patterns, operating conditions, manufacturing variability, or other factors.

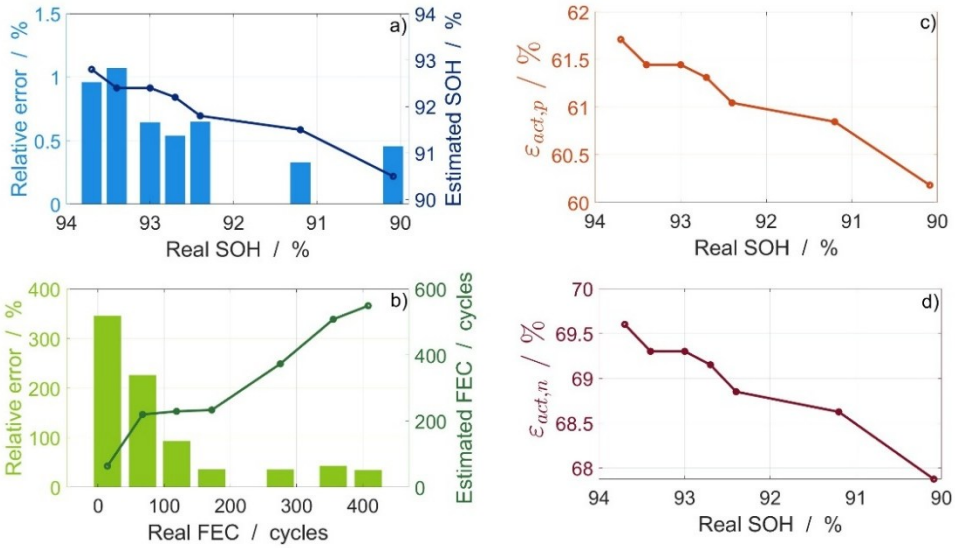

Figure 6: DDM results: a) and b) show the SOC and FEC estimates vs laboratory measurements, respectively, showing relative error (left y-axis) and estimated values (right y-axis). c) and d) show the active material volume fractions for the positive and negative electrodes, respectively.

The DDM also estimates $\varepsilon_{act,p}$ and $\varepsilon_{act,n}$, shown in ***Figure 6 c)*** and ***d)***, both of which decrease uniformly with SOH.

Once this information is obtained, the HM proceeds to the second phase, estimating the future performance of the SLB under a LEV scenario.

The PBM forecasts performance by repeating the desired usage pattern, illustrated in Figure 5, which reaches its maximum depth of discharge between hours 16 and 17, corresponding to a total discharge of 3.9 Ah. Assuming the scooter is fully recharged after each use, the LG-M50, with a nominal capacity of 5 Ah, remains suitable for this second-life application until reaching 78% SOH. Based on this assumption, the PBM estimates the RUL at different SOH levels by modelling the SOH evolution for the case study. ***Table 3*** summarizes the HM results, indicating that the SLB would remain operational for 5.3 to 4.4 years, depending on the initial SOH of the LIB.

Table 3: RUL estimated by the PBM at different SOH initial states

SOH measured / %	RUL / years
93.7	5.3
93.4	5.2
93	5.2
92.7	5
92.4	4.8
91.2	4.7
90.1	4.4

6. CONCLUSIONS

This study demonstrates the potential of HMs that combine DDMs with PBMs to evaluate and forecast the performance of SLBs. A 24-hour scooter usage profile, measured under real-world operating conditions, was used to validate the model. The DDM achieved high SOH estimation accuracy, with deviations of 0.3–1.1% from experimental data. In contrast, FEC predictions showed larger errors, likely due to their higher sensitivity to prior usage, operating conditions, manufacturing variability, and the limited training dataset of only six cells. Differences in degradation patterns between the tested SLB cell and the training cells may also have contributed to the substantial deviation in FEC estimation.

The study also highlights the value of active material volume fraction as a key parameter for linking DDMs and PBMs. This parameter enhances degradation modelling, provides insight into electrode-level changes, and improves the adaptability of PBMs for already degraded batteries. PBMs proved useful for predicting future performance, including SOH evolution and RUL, relevant for LEV applications.

The reduced number of cells considered in this study may not capture the full range of degradation patterns or effects such as manufacturing variability, limiting the generalizability of the results. Future work will expand the dataset to include more LIBs with different usage histories, test the HM in more conditions and study how certain degradation mechanisms affect the $\varepsilon_{act,k}$ of each electrode. Overall, the proposed HM effectively integrates the strengths of DDMs and PBMs, offering a promising tool for promoting SLB use in sustainable energy storage systems while reducing testing time. Although demonstrated on NMC811 cells, the methodology is adaptable to other chemistries and applications, with practical value for research, industry, and areas like warranty design.

ACKNOWLEDGEMENTS

This research has been conducted under the frame of the projects FREE4LIB and iBattMan, both funded by the European Union's program Horizon Europe under the grants No 101069890 and 101138856. Also, by Generalitat de Catalunya under the CERCA Programme, in the ACCIO program under the project DIGIBAT and the Joan Oró predoctoral, co-funded by the European Social Fund Pluss (2024 FI-1 00324). Finally, by the project "Optimal hybrid AI strategies for battery management and state estimation", via the oc2-2024-

TES-02 issued and implemented by the ENFIELD project, under the grant agreement No 101120657.

REFERENCES

Chen, C.-H., Planella, F. B., O'Regan, K., Gastol, D., Widanage, W. D., & Kendrick, E. (2020). Development of experimental techniques for parameterization of multi-scale lithium-ion battery models. Journal of The Electrochemical Society, 167(8), 080534. https://doi.org/10.1149/1945-7111/ab9050

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '16) (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Huotari, M., Arora, S., Malhi, A., & Främling, K. (2021). Comparing seven methods for state-of-health time series prediction for the lithium-ion battery packs of forklifts. Applied Soft Computing, 111, 107670. https://doi.org/10.1016/j.asoc.2021.107670

Jokar, A., Rajabloo, B., Désilets, M., & Lacroix, M. (2016). Review of simplified pseudo-two-dimensional models of lithium-ion batteries. Journal of Power Sources, 327, 44–55. https://doi.org/10.1016/j.jpowsour.2016.07.036

Navidi, S., Thelen, A., Li, T., & Hu, C. (2024). Physics-informed machine learning for battery degradation diagnostics: A comparison of state-of-the-art methods. Energy Storage Materials, 68, 103343. https://doi.org/10.1016/j.ensm.2024.103343

Nitta, N., Wu, F., Lee, J. T., & Yushin, G. (2015). Li-ion battery materials: Present and future. Materials Today, 18(5), 252–264. https://doi.org/10.1016/j.mattod.2014.10.040

O'Kane, S. E. J., Ai, W., Madabattula, G., Alonso-Alvarez, D., Timms, R., Sulzer, V., Edge, J. S., Wu, B., Offer, G. J., & Marinescu, M. (2022). Lithium-ion battery degradation: How to model it. Physical Chemistry Chemical Physics, 24, 7909–7922. https://doi.org/10.1039/D2CP00417H

Sulzer, V., Marquis, S. G., Timms, R., Robinson, M., & Chapman, S. J. (2021). Python battery mathematical modelling (PyBaMM). Journal of Open Research Software, 9(4). https://doi.org/10.5334/jors.309

Von Bülow, F., Heinrich, F., & Paxton, W. A. (2024). The future of battery data and the state of health of lithium-ion batteries in automotive applications. Communications Engineering, 3, 173. https://doi.org/10.1038/s44172-024-00299-w

Wang, J., Peng, Q., Meng, J., Liu, T., Peng, J., & Teodorescu, R. (2024). A physics-informed neural network approach to parameter estimation of lithium-ion battery electrochemical model. Journal of Power Sources, 621, 235271. https://doi.org/10.1016/j.jpowsour.2024.235271