Adaptive Health Monitoring of Second-Life Batteries

Xiaofan Cui[1], *Member, IEEE*, Muhammad Aadil Khan[1], *Student Member, IEEE*, Surinder Singh[2], Ratnesh Sharma[2], and Simona Onori[1], *Senior Member, IEEE*

***Abstract*—As surplus of second-life (SL) batteries becomes available, one critical obstacle between their widespread adoption is the accurate estimation and monitoring of their state-of-health (SOH). Various retired battery packs with a lack of knowledge of their historical usage are used to set up new SL battery systems. The paper presents a new online adaptive health estimation method designed to address these practical challenges associated with SL battery systems. This method utilizes the real-time data obtained from the SL batteries used for grid energy storage systems. The proposed approach is validated on a laboratory-aged experimental data set of retired EV batteries. Through dynamic adjustment of estimator gains, the approach can effectively accommodate the unique characteristics of individual cells, enhancing its adaptability and robustness. We prove theoretically that this method has bounded-input-bounded-output (BIBO) stability.**

I. Introduction

In the effort to reduce the impact of climate change, transition towards sustainable energy is an important factor. While all major automotive manufacturers have started including electric vehicles (EVs) in their fleet [1], [2], some manufacturers have completely moved away from internal combustion engine vehicles [3]. As the demand for lithium-ion batteries (LIBs) increase, it also increases the pressure on the supply chain for the raw materials needed to produce these batteries [4]. One opportunity to overcome this challenge is to reuse retired EV batteries after their first-life (FL). The availability of approximately 70-80% of the nominal fresh cell capacity of these batteries makes them immensely suitable for stationary applications such as power generation, grid storage, and residential services.

Second-life (SL) batteries, although promising, pose a number of challenges for effective usage. Due to the fact that SL batteries are build up from various different battery packs, accurately estimating their state-of-health (SOH) is difficult since different batteries have experienced different usage conditions resulting in various different degradation behaviors. In order to safely use SL batteries, a dedicated battery management system (BMS_2) is proposed, which specifically focuses on the health estimation problem of retired batteries, and provides a framework to mitigate some of the challenges with using these batteries.

[1]Xiaofan Cui, Muhammad Aadil Khan, and Simona Onori are with the Department of Energy Science & Engineering, Stanford University, CA 94305 cuixf@seas.ucla.edu, maadilk@stanford.edu, sonori@stanford.edu (corresponding author). Xiaofan Cui is currently a faculty at Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA, 90095, USA.

[2]Surinder Singh and Ratnesh Sharma are with the Relyion Energy, Santa Clara, CA, 95054, USA.

Notations: To establish the theoretical framework, the following notations are used in this paper:

(1) $\|\cdot\|_2$ represents the $\mathcal{L}_2$ norm. The $\mathcal{L}_2$ norm of a sequence $\{x_n\} = (x_1, x_2, \cdots, x_N)$ is captured by the formula $\|\{x_n\}\|_2 = \sqrt{\sum_{n=1}^{N} x_n^2}$.

(2) $\|\cdot\|$ represents the $\mathcal{L}_\infty$ norm. The $\mathcal{L}_\infty$ norm of a sequence $\{x_n\} = (x_1, x_2, \cdots, x_N)$ is captured by the formula $\|\{x_n\}\|_\infty = \max_{n=1}^{N} |x_n|$.

(3) $\mathbb{R}^+ = \{z \in \mathbb{R} : z > 0\}$.

(4) Sequence notation $\{a_n\}_{n\leq N} \triangleq \{a_1, a_2, \cdots, a_N\}$. a_n represents the general item of the sequence $\{a_n\}_{n\leq N}$. For ease of notation, if not specified, $\{a_n\}$ is equivalent to $\{a_n\}_{n\leq N}$.

(5) Three performance metrics include Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Root Mean Squared Percentage Error (RMSPE), which are defined by

$$\text{MAPE} = \frac{1}{M} \sum_{y\in Y,\ \hat{y}\in\hat{Y}} \frac{|\hat{y} - y|}{y} \times 100\,\% \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{y\in Y,\ \hat{y}\in\hat{Y}} (\hat{y} - y)^2} \times 100\,\% \tag{2}$$

$$\text{RMSPE} = \sqrt{\frac{1}{M} \sum_{y\in Y,\ \hat{y}\in\hat{Y}} \left(\frac{\hat{y} - y}{y}\right)^2} \times 100\,\% \tag{3}$$

where Y is the measured data, $\hat{Y}$ is the model estimation, and M is the number of samples.

Several approaches have been explored for SOH estimation, including semi-empirical methods, physics-based models, and data-driven approaches [5]. While semi-empirical methods exhibit strong performance under controlled laboratory conditions, their accuracy degrades when applied to real-world scenarios [5]. On the other hand, physics-based models, relying on partial differential equations, offer detailed insights into internal battery parameters but they are computationally demanding and require the identification of numerous parameters [6]. In contrast, data-driven models leverage available data to predict target output behavior without explicitly modeling underlying dynamics. These models offer computational efficiency during testing, making them appealing for online applications. SOH estimation using data-driven models often employs regression techniques like Linear Regression [7], Support Vector Regression [8], and Gaussian Process Regression [9]. Additionally, neural networks have been utilized in several studies [10], [11]. However,

achieving optimal performance with neural networks necessitates large datasets and extensive hyperparameter tuning.

Although data-driven techniques have demonstrated to be effective, current approaches to adaptive estimation rely on periodically-measured health indicators obtained through reference performance tests (RPTs), with feedback used to update the model [12], [13]. However, this method is impractical for BMS_2 due to the inability to continuously disengage SL batteries once deployed [14]. An alternative popular strategy involves transfer learning, where a baseline model is initially trained using a neural network. Subsequently, the initial layers of the model are fixed, while the final layers are retrained using data from the target domain [15], [16]. Nevertheless, the transfer learning is limited by the target domain used to re-train the model. In practical field operations, incoming data often exhibit high uncertainty, posing challenges in ensuring consistent model performance.

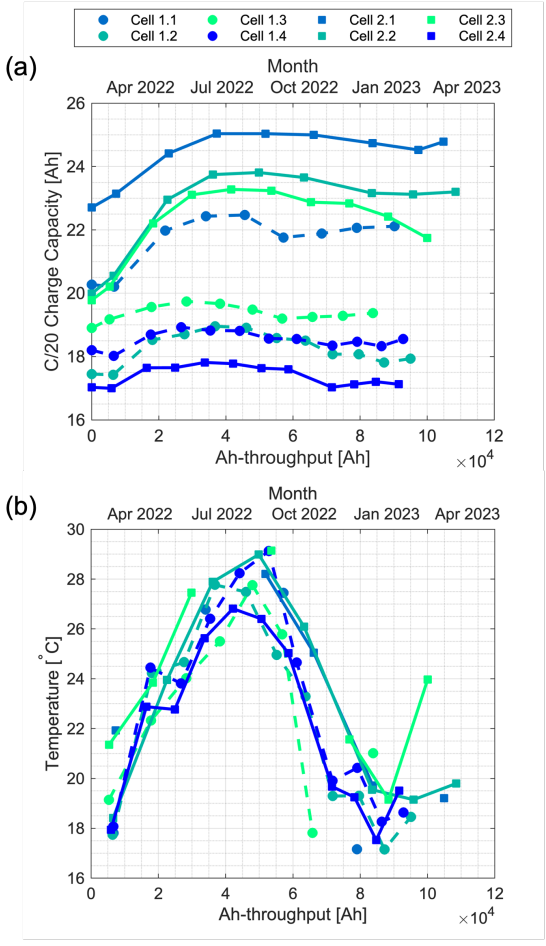

Fig. 1: The capacity and temperature data for all cells are depicted. Graph (a) illustrates the C/20 charge capacity plotted against Ah-throughput, demonstrating the increasing capacity as the cells undergo cycling. Graph (b) shows temperature plotted against Ah-throughput, displaying a parabolic shape with peak values typically observed during August/September 2022, indicative of the seasonal temperature variation's impact on all cells. Discontinuous lines indicate missing data points for certain cells.

Designing the health estimation task for BMS_2 requires not only the online functionality but also adaptability to variation in input signals. This involves the continuous adjustment and evolution of the estimator model as more online measurements become available. Considering the auto-correlation of SOH time series [17], historical health data becomes crucial for accurately estimating the current SOH. Moreover, it is practically significant for health estimators to converge within specific error bounds while provable mathematical guarantees.

To address the uncertainty inherent in SOH estimation algorithms trained offline, this study introduces a clustering-based adaptive estimation method. Given that adaptive SOH estimation relies on the real-time flow of incoming data, it is crucial to ensure the boundedness of model estimation. Accordingly, we theoretically establish the stability of the adaptive framework using bounded-input bounded-output (BIBO) stability analysis. We compare the outcomes derived from the adaptive framework with those from the offline model to demonstrate the enhancement in estimation accuracy resulting from our model adaptation.

II. Retired Battery Dataset

For this work, the dataset consists of eight retired pouch cells obtained from EV battery packs with LMO/graphite chemistry. These cells operate within the voltage window from 2.5 V to 4.2 V with the nominal capacity of 32.5 Ah. The cells are present inside a Heating, Ventilation and Cooling (HVAC) controlled room. Although, the temperature of the room is maintained at a human-comfort level, the seasonal variations in temperature influence the cell surface temperature as shown in Fig. 1(b).

The proposed experimental campaign consists of aging cycles and three different reference performance tests (RPTs). The aging-cycle current profile is structured to replicate the grid-energy-storage scenario. The RPTs include a C/20 capacity test, a Hybrid Pulse Power Characterization (HPPC) test, and a C/40 Open-Circuit Voltage (OCV) characterization [18]. It is worth noting that all the tests are executed within the voltage range of 3 V to 4 V, except for the OCV test, which is performed within the voltage range of 2.5 V to 4.2 V. Such voltage derating strategies have been observed to decelerate the degradation process within cells [19], making them particularly suitable for SL applications. Comprehensive details regarding the experimental campaign, including its duration and the formulation of testing protocols, are elaborated upon in [20].

The C/20 discharge capacity trajectory is shown in Fig. 1(a). In contrast to conventional battery degradation profiles, these cells exhibit an increase in capacity. Detailed explanations are provided in [20], attributing this behavior to the fluctuating temperatures experienced by the cells during the experimental campaign. This dataset emphasizes the temperature-dependent behavior of SL batteries in grid energy storage applications, where strict temperature control is impractical. Additionally, the positive impact of voltage

derating on battery health is observable, as cells, upon reaching the one-year mark from the start of testing, still maintain capacities equal to or higher than their initial capacities.

III. Offline Data-driven Model

The selection of a data-driven model depends on various factors including data complexity, available computational resources, and training time. Considering these criteria, we opt for an Elastic-Net Regression (ENR) model. This choice offers simplicity in training and adaptability, along with minimal computational demands. To mitigate overfitting caused by the dataset's limited size, the loss function of the model comprises of a linear combination of L_1 (lasso) and L_2 (ridge) regularization terms. The optimization problem addressed during the training phase can be formulated as follows

$$\hat{\beta} = \underset{\beta_0,\beta,\lambda}{\arg\min} \ \|Y - X\beta - \beta_0 \mathbf{1}_{n\times 1}\|_2 + \lambda(1-\alpha)\|\beta\|_2^2 + \lambda\alpha\|\beta\|_1. \quad (4)$$

Here, $Y \in \mathbb{R}^n$ signifies the SOH indicators, $X \in \mathbb{R}^{n\times m}$ encompasses m features each with n observations, $\beta \in \mathbb{R}^m$ consists of m regression coefficients, $\beta_0 \in \mathbb{R}$ denotes the scalar intercept, and $\lambda, \alpha \in \mathbb{R}$ serve as hyperparameters to modulate the relative impact of L_1 and L_2 regularization terms.

A. Offline ENR

The SOH indicator adopted in this study is defined as the C/20 charge capacity. In the dataset under study, it is observed that the disparity between charge and discharge capacities is negligible. Out of a total of 8 cells, 6 cells are allocated for training purposes, while the remaining 2 cells are designated for testing. The term "offline" is used to describe the model because the estimation outcomes are solely derived from the existing training dataset without any efforts to adapt the model to enhance its estimation performance based on new data. Considering a total of 8 cells, the count of the possible two cell combinations amounts to 28. The evaluation of the offline ENR model's performance is conducted through metrics including RMSE, RMSPE, and MAPE.

Table I illustrates the performance of the offline ENR model across 28 distinct combinations of test sets. For test set (1.2, 2.1), the model achieves the best performance with an RMSE of 0.28 Ah, RMSPE of 1.21% and MAPE of 1%. Out of all the test sets, only (1.1, 2.3) and (2.2, 2.3) have an RMSE greater than 1 Ah, and both sets contain Cell 2.3. As mentioned in our work in [20], Cell 2.3 experienced technical issues during data collection leading to presence of anomalous data for this cell. Consequently, if Cell 2.3 is not a part of the training set, the model performance deteriorates resulting in poor performance for capacity estimation. This limitation of the offline model also highlights the importance of having an online adaptive model, which can adapt to the uncertainties in the incoming data, and provide consistent estimation results within specific error bounds.

TABLE I: Performance of offline ENR model

Test set	RMSE [Ah]	RMSPE [%]	MAPE [%]
(1.1, 1.2)	0.495699	2.451021	1.565399
(1.1, 1.3)	0.555976	2.769582	2.167772
(1.1, 1.4)	0.484275	2.394690	1.602038
(1.1, 2.1)	0.628837	2.920140	2.046725
(1.1, 2.2)	0.682332	3.193374	2.437554
(1.1, 2.3)	1.495901	6.595751	5.265551
(1.1, 2.4)	0.566506	2.915141	2.185236
(1.2, 1.3)	0.689619	3.713645	3.319282
(1.2, 1.4)	0.396715	2.147421	1.957864
(1.2, 2.1)	**0.280542**	**1.218656**	**1.006326**
(1.2, 2.2)	0.529458	2.538446	2.010231
(1.2, 2.3)	0.800931	3.533838	2.553109
(1.2, 2.4)	0.485609	2.734561	2.507264
(1.3, 1.4)	0.355872	1.892205	1.683591
(1.3, 2.1)	0.562813	2.580422	2.353598
(1.3, 2.2)	0.608824	2.832768	2.695224
(1.3, 2.3)	0.763585	2.400556	2.894587
(1.3, 2.4)	0.457407	2.533389	2.187910
(1.4, 2.1)	0.401952	1.827907	1.246798
(1.4, 2.2)	0.530269	2.527535	1.905807
(1.4, 2.3)	0.791817	3.497391	2.661257
(1.4, 2.4)	0.323652	1.802888	1.377056
(2.1, 2.2)	0.737153	3.225221	2.668015
(2.1, 2.3)	0.704329	3.105652	2.436052
(2.1, 2.4)	0.444313	2.148165	1.700114
(2.2, 2.3)	1.038041	4.536129	4.018927
(2.2, 2.4)	0.571021	3.012412	2.415986
(2.3, 2.4)	0.672987	3.190073	2.869701

IV. Adaptive Estimator

The health estimation of BMS_2 necessitates not only online functionality but also adaptiveness, implying that the estimator model should adjust and refine itself with the accumulation of more online measurements.

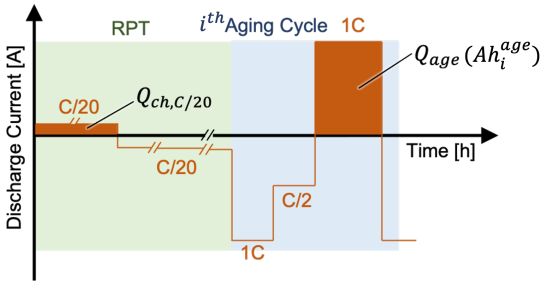

Fig. 2: Graphical representation of aging-cycle charge throughput, Q_{age}, and C/20 charge capacity, $Q_{ch,C/20}$. Positive current indicates charge and negative current indicates discharge.

Definition 1. Given the monotonically increasing accumulated Ah-throughput sequence $\{Ah_n^{age}\} = (Ah_1^{age}, Ah_2^{age}, \cdots, Ah_N^{age})$, $Ah_i^{age} \in \mathbb{R}^+$ and the aging-cycle charge throughput vector $Q_{age} \in \mathbb{R}^N$, then the aging-cycle charge throughput trajectory is a curve denoted by $Q_{age}(\{Ah_n^{age}\})$.

Remark: Q_{age} is calculated as shown in Fig. 2. While Q_{age} is one of the numerous online-available features, it stands out as the most relevant feature with respect to our SOH indicator $Q_{ch,C/20}$ [20]. Consequently, our adaptive estimation algorithm utilizes this feature to update the model.

Definition 2. Given the monotonically increasing accumulated Ah-throughput sequence $\{Ah_n^{ch,c/20}\} = \left(0, Ah_2^{ch,c/20}, \cdots, Ah_N^{ch,c/20}\right)$, $Ah_i^{ch,c/20} \in \mathbb{R}^+$ and the C/20 charge capacity vector $Q_{ch,C/20} \in \mathbb{R}^N$, the C/20 charge capacity trajectory is a curve denoted by $Q_{ch,C/20}(\{Ah_n^{ch,c/20}\})$.

Remark: $Q_{ch,C/20}$ is calculated as shown in Fig. 2.

The *normalized SOH indicator* we use is the normalized C/20 charge capacity, calculated by

$$\bar{Q}(Ah_n) = \frac{Q_{ch,C/20}(Ah_n)}{Q_{ch,C/20}(0)}, \tag{5}$$

where $Q_{ch,C/20}(0)$ denotes the $C/20$ charge capacity of the cell measured at the beginning of the second life, by following the similar definition in (2), the *normalized SOH indicator trajectory* can be defined.

Definition 3. *The training set* is denoted by indices $1, \cdots, K$, where K is the number of cells in the training set.

Definition 4. *The test set* is denoted by index z.

Definition 5. *The aging-cycle charge training trajectory set* is denoted by $Q_{age}^1, \cdots, Q_{age}^K$, where K, the number of cells, is the cardinality of the set.

Similarly, the *C/20 charge capacity training trajectory* $Q_{ch,C/20}^{train}$ is denoted by $Q_{ch,C/20}^1, \cdots, Q_{ch,C/20}^K$ and the *normalized SOH training trajectory* $\bar{Q}^{train}$ is denoted by $\bar{Q}^1, \cdots, \bar{Q}^K$, respectively.

Definition 6. The *data-driven battery health estimator* $\mathcal{E}_H$ is a type of data-driven estimator with:

$$\hat{Q}_{ch,C/20}^{test} = \mathcal{E}_H\left(Q_{age}^{test}; Q_{age}^{train}, Q_{ch,C/20}^{train}\right) \tag{6}$$

$$e_n = \hat{Q}_{ch,C/20}^{test} - Q_{ch,C/20}^{test} \tag{7}$$

where $Q_{ch,C/20}^{train}$ and $Q_{ch,C/20}^{test}$ are the C/20 charge capacity trajectories in Definition 2. Q_{age}^{train} and Q_{age}^{test} are the aging-cycle charge throughput trajectories in Definition 1.

Stability holds practical significance for adaptive estimation laws, ensuring that the proposed estimation process remains robust and does not lead to divergence. The stability notion employed by this paper is as following:

Definition 7. A data-driven adaptive estimator $\mathcal{E}_H$ defined by Definition 6 is *BIBO stable* if for any bounded input signal with $\|Q_{age}^{train}\| < \infty$, $\|Q_{age}^{test}\| < \infty$, $\|Q_{ch,C/20}^{train}\| < \infty$, $\|Q_{ch,C/20}^{test}\| < \infty$, the error e_n satisfies $\|e_n\| < \infty$.

A. Clustering-based adaptive estimation

The primary objective of the clustering-based adaptive estimation algorithm is to select the trajectory within the training set that closely aligns with the trajectory in the test set. This alignment is assessed using the following distance function:

Definition 8. The *distance* between two aging-cycle charge throughput trajectories Q^x and Q^y are

$$\text{dist}(Q^x, Q^y) = \sqrt{\sum_{i=1}^{N} \left(Q^x(Ah_i)) - Q^y(Ah_i))\right)^2}, \tag{8}$$

Remark: The distance function $\text{dist}(Q^x, Q^y)$ is non-negative and symmetry owing to the characteristics of the $\mathcal{L}_2$ norm. This distance measure can be easily expanded to evaluate the disparity between other trajectories, such as the C/20 charge capacity trajectory, among others. Upon acquiring a new sample of cell z, the classification is updated. The resulting *Classification index sequence* is defined as:

Definition 9. *Classification index sequence* for cell z $\{S_n^z\} = \left(S_1^z, S_2^z, \cdots, S_N^z\right)$ is a discrete-time sequence defined by:

$$S_n^z = \underset{1\le k\le K}{\text{argmin}}\ \text{dist}\left(Q_{age}^z(\{Ah_m\}_{m\le n}), Q_{age}^k(\{Ah_m\}_{m\le n})\right), \tag{9}$$

where Q_{age}^z and Q_{age}^k are the aging-cycle charge trajectories of cell z and k, respectively. The distance function $dist$ has been defined in Definition 8.

Theorem 1. Consider a data-driven estimator described in Definition 6. Let the battery cells 1 to cell K be the training set and cell z be the test set. If the model parameters are adapted, with the tunning parameters λ_k according to

$$\lambda_k = \frac{\sum_{n=1}^{N} \chi(S_n^z = k) Ah_n}{\sum_{n=1}^{N} Ah_n}, \tag{10}$$

where

$$\chi(S_n^z = k) = \begin{cases} 1 & \text{if } I_n^z = k, 1 \le k \le K \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

I_n^z is defined in (9). Then the adaptive estimator, formulated as

$$\hat{Q}_{ch,C/20}^z(Ah_n) = Q_{ch,C/20}^z(0)\hat{\bar{Q}}^z(Ah_n), \tag{12}$$

$$\hat{\bar{Q}}^z(Ah_n) = \sum_{k=1}^{K} \lambda_k \bar{Q}^k(Ah_n), \tag{13}$$

where $\hat{Q}_{ch,C/20}^z$, the estimation target, represents the C/20 charge capacity of cell z, is BIBO stable.

Proof. From (10), $\{\lambda_k\}$ has the following properties:

$$\sum_{k=1}^{K} \lambda_k = 1, \tag{14}$$

$$0 \leq \lambda_k \leq 1, \quad \text{for } 1 \leq k \leq K. \tag{15}$$

Property (14) can be proved by:

$$\sum_{k=1}^{K} \lambda_k = \sum_{k=1}^{K} \frac{\sum_{n=1}^{N} \chi(S_n^z = k) Ah_n}{\sum_{n=1}^{N} Ah_n} \tag{16}$$

$$= \sum_{n=1}^{N} \frac{\sum_{k=1}^{K} \chi(S_n^z = k) Ah_n}{\sum_{n=1}^{N} Ah_n} \tag{17}$$

$$= \sum_{n=1}^{N} \frac{Ah_n}{\sum_{n=1}^{N} Ah_n} = 1. \tag{18}$$

Because of (14) and λ_k is positive, Property (14) can be easily verified.

For ease of notation, in this proof, the trajectory variable $Q_{ch,C/20}$ is denoted by Q and the scalar variable $Q_{ch,C/20}(0)$ is denoted by Q_{bsl}. From the BIBO stability Definition (7), the training response trajectory is bounded:

$$\|Q^k\| < \infty, \qquad \text{for } 1 \leq k \leq K. \tag{19}$$

The test response trajectory is bounded:

$$\|Q^z\| < \infty. \tag{20}$$

The error trajectory between the estimated capacity and the true capacity follows:

$$\left\|\hat{Q}^z - Q^z\right\| \overset{(5)}{=} \left\| Q_{bsl}^z \sum_{k=1}^{K} \lambda_k \bar{Q}^k - Q^z \right\| \tag{21}$$

$$\overset{(14)}{=} \left\| Q_{bsl}^z \sum_{k=1}^{K} \lambda_k \bar{Q}^k - \sum_{k=1}^{K} \lambda_k Q^z \right\| \tag{22}$$

$$\overset{(5)}{=} \left\| Q_{bsl}^z \sum_{k=1}^{K} \lambda_k \bar{Q}^k - Q_{bsl}^z \sum_{k=1}^{K} \lambda_k \bar{Q}^z \right\| \tag{23}$$

$$= Q_{bsl}^z \left\| \sum_{k=1}^{K} \lambda_k \left(\bar{Q}^k - \bar{Q}^z \right) \right\|. \tag{24}$$

From (15) and (24)

$$\left\|\hat{Q}^z - Q^z\right\| \leq Q_{bsl}^z \max_{1 \leq k \leq K} \{\left\|\bar{Q}^k - \bar{Q}^z\right\|\} \tag{25}$$

$$\leq \max_{1 \leq k \leq K} \{\left\|Q^k\right\|\} + \|Q^z\| < \infty. \tag{26}$$

Therefore, the error trajectory is bounded. From Definition 7, the BIBO stability is guaranteed. □

In all, the adaptive law employed in the estimation process ensures that the estimation remains stable, and the maximum error in health estimation can be constrained. Moreover, adjusting the parameter λ allows for further reduction of the error bound.

V. Results and Discussion

The proposed algorithm's performance is showcased through a case study, wherein Cell 1.4 is reserved for testing while the models are trained on the remaining 7 cells. The C/20 charge capacity trajectories estimated by the ENR model, online clustering-based estimation model, and actual capacity are depicted in Fig. 3(a).

Initially, the clustering index is set at 2, corresponding to Cell 1.2. As the Ah-throughput surpasses 2×10^4 Ah, the classification index converges to the correct cluster (cluster 8 / Cell 2.4) as shown in Fig. 3(c). However, beyond an Ah-throughput of 6.5×10^4 Ah, the classification index fluctuates to the incorrect cluster (cluster 2 / Cell 1.2).

The adaptive SOH estimator continually processes real-time data influx, resulting in lower estimation errors than the offline method. Fig. 3(b) illustrates that using solely the ENR model leads to a maximum absolute pointwise percentage error of 3.202 % and a root mean squared pointwise percentage error of 1.8166 %. Conversely, employing the clustering-based online estimation model reduces the maximum pointwise percentage error to 3.1532 %, and the root mean squared pointwise percentage error to 1.6676 %.

This example underscores the significance of weight adaptation. The weight parameter λ balances the influence of the current-step classification result with historical-step classification results, which are retained by summing their corresponding Ah-throughputs. This mechanism effectively mitigates the online clustering-based estimation method's high sensitivity to the current-step classification result.

VI. Conclusions

Due to its flexibility and model agnosticism, data-driven health estimation is an emerging approach for evaluating the health of SL batteries. In order to facilitate the performance of SL battery energy storage systems in real-world settings, we introduce an online adaptive data-driven health estimation technique with the stability guarantee. This method has been validated using a dataset comprised of lab-aged second-life batteries retired from commercial electric vehicles. Our proposed method demonstrates the potential in minimizing estimation errors when compared to conventional offline health estimation methods.

ACKNOWLEDGMENT

The authors would like to thank Surinder Singh and Ratnesh Sharma for collecting the data described in [20] and StorageX Initiative within the Precourt Institute of Energy at Stanford University for the financial support.

References

[1] N. Lutsey, M. Grant, S. Wappelhorst, and H. Zhou, "Power play: How governments are spurring the electric vehicle industry." ICCT Washington, DC, USA, 2018.

[2] J. Eddy, A. Pfeiffer, and J. van de Staaij, "Recharging economies: The ev-battery manufacturing outlook for europe," *McKinsey & Company*, 2019.

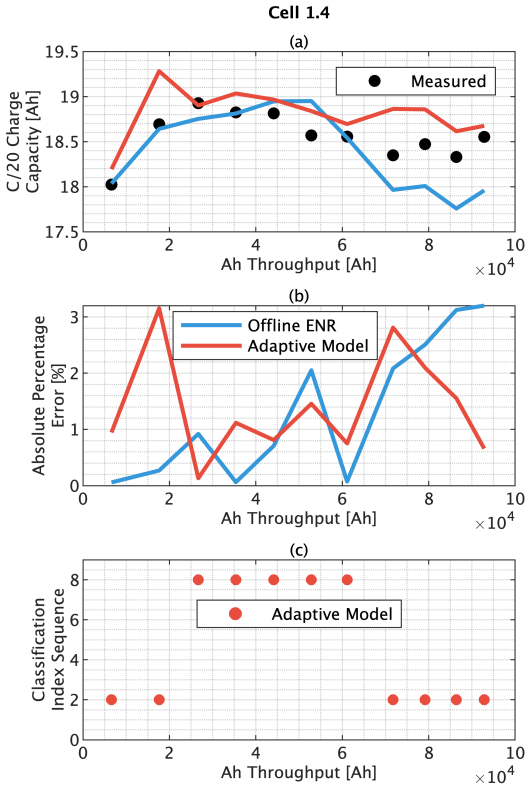

Fig. 3: Evaluation of the performance of the online clustering-based estimator for cell 1.4. (a) A comparison between the C/20 charge capacity trajectories estimated from the offline Elastic-Net Regression (ENR) and the online clustering-based model against the measured capacity. (b) Comparison of the pointwise absolute capacity estimation percentage error of the offline ENR with that of the online clustering-based model. The Root Mean Square Error (RMSE) of the online clustering-based model stands at 1.6676 %. (c) Presentation of the classification index sequence, as defined in Definition 9, of the online clustering-based estimator.

[3] P. Pavlínek, "Transition of the automotive industry towards electric vehicle production in the east european integrated periphery," *Empirica*, vol. 50, no. 1, pp. 35–73, 2023.

[4] B. Jones, V. Nguyen-Tien, and R. J. Elliott, "The electric vehicle revolution: Critical material supply chains, trade and development," *The World Economy*, vol. 46, no. 1, pp. 2–26, 2023.

[5] X. Hu, X. Deng, F. Wang, Z. Deng, X. Lin, R. Teodorescu, and M. G. Pecht, "A Review of Second-Life Lithium-Ion Batteries for Stationary Energy Storage Applications," *Proceedings of the IEEE*, vol. 110, no. 6, pp. 735–753, jun 2022.

[6] G. Pozzato, S. B. Lee, and S. Onori, "Modeling degradation of Lithium-ion batteries for second-life applications: preliminary results," *CCTA 2021 - 5th IEEE Conference on Control Technology and Applications*, pp. 826–831, 2021.

[7] Y. Jiang, J. Jiang, C. Zhang, W. Zhang, Y. Gao, and N. Li, "State of health estimation of second-life lifepo4 batteries for energy storage applications," *Journal of Cleaner Production*, vol. 205, pp. 754–762, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652618328725

[8] J. Wei, G. Dong, and Z. Chen, "Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5634–5643, 2018.

[9] A. Takahashi, A. Allam, and S. Onori, "Evaluating the feasibility of batteries for second-life applications using machine learning," *iScience*, vol. 26, no. 4, p. 106547, 2023. [Online]. Available: https://doi.org/10.1016/j.isci.2023.106547

[10] C. Zhang, J. Jiang, W. Zhang, Y. Wang, S. M. Sharkh, and R. Xiong, "A novel data-driven fast capacity estimation of spent electric vehicle lithium-ion batteries," *Energies*, vol. 7, no. 12, pp. 8076–8094, 2014. [Online]. Available: https://www.mdpi.com/1996-1073/7/12/8076

[11] A. Bhatt, W. Ongsakul, N. Madhu, and J. G. Singh, "Machine learning-based approach for useful capacity prediction of second-life batteries employing appropriate input selection," *International Journal of Energy Research*, vol. 45, no. 15, pp. 21 023–21 049, dec 2021.

[12] Y. Zhang, T. Wik, J. Bergström, M. Pecht, and C. Zou, "A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data," *Journal of Power Sources*, vol. 526, p. 231110, 2022.

[13] C. She, Y. Li, C. Zou, T. Wik, Z. Wang, and F. Sun, "Offline and online blended machine learning for lithium-ion battery health state estimation," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 2, pp. 1604–1618, 2021.

[14] Y. Zhang, T. Wik, J. Bergström, M. Pecht, and C. Zou, "A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data," *Journal of Power Sources*, vol. 526, apr 2022.

[15] K. Liu, Q. Peng, Y. Che, Y. Zheng, K. Li, R. Teodorescu, D. Widanage, and A. Barai, "Transfer learning for battery smarter state estimation and ageing prognostics: Recent progress, challenges, and prospects," *Advances in Applied Energy*, p. 100117, 2022.

[16] J. Lu, R. Xiong, J. Tian, C. Wang, and F. Sun, "Deep learning to estimate lithium-ion battery state of health without additional degradation experiments," *Nature Communications*, vol. 14, no. 1, p. 2760, 2023.

[17] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis *et al.*, "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, no. 5, pp. 383–391, 2019.

[18] S. Ha, G. Pozzato, and S. Onori, "Electrochemical characterization tools for lithium-ion batteries," *Journal of Solid State Electrochemistry*, vol. 28, no. 3, pp. 1131–1157, 2024. [Online]. Available: https://doi.org/10.1007/s10008-023-05717-1

[19] C. P. Aiken, E. R. Logan, A. Eldesoky, H. Hebecker, J. Oxner, J. Harlow, M. Metzger, and J. Dahn, "Li[$Ni_{0.5}Mn_{0.3}Co_{0.2}$] O_2 as a superior alternative to $LiFePO_4$ for long-lived low voltage li-ion cells," *Journal of The Electrochemical Society*, vol. 169, no. 5, p. 050512, 2022.

[20] X. Cui, M. A. Khan, G. Pozzato, R. Sharma, S. Singh, and S. Onori, "Taking second-life batteries from exhausted to empowered using experiments, data analysis, and health estimation," *Cell Reports Physical Science*, 2024.