Data-selection for state estimation of large-scale battery systems

Zhuo Wang*, Daniel T. Gladwin, Matthew J. Smith, Thomas L. Fantham
Department of Electronic and Electrical Engineering
Sheffield, United Kingdom
Email: zhuowang@sheffield.ac.uk, d.gladwin@sheffield.ac.uk, matt.j.smith@sheffield.ac.uk, tfantham1@sheffield.ac.uk

***Abstract*—Large-scale battery energy storage systems (BESS) are drawing the attention of researchers as numbers installed globally is rising rapidly. Like single cells, the battery states, most importantly state of charge (SOC) and state of health (SOH) of BESSs are essential for their operation. However, for large-scale battery systems, the data granularity, accuracy and quality are limited compared with the cell-level. To achieve accurate state estimation of battery systems the selection of data used for processing is essential. In this paper, it is shown that how to evaluate and select system-level data for SOC and SOH estimation. These methods are expected to be used for other BESSs.**

***Index Terms*—Battery System, State of Charge, State of health, Data selection, Invalid Data.**

I. Introduction

The number of electrical grid-connected Battery energy storage systems (BESS) has been growing significantly over the last 5 years and play a crucial role in managing the electrical grid where more renewable resources generate clean, but variable, electricity. State estimation, mainly state of charge (SOC) and state of health (SOH) of a BESS are critical for its operations.

The authors of this paper have proposed using Kalman filter (KF) methods, and a total least-square method (TLS) to estimate system SOC and capacity of large-scale BESS using system-level data in [1]. In this paper these algorithms are demonstrated using a real-world 2MW/1MWh grid-connected battery. Applying cell-level state estimation techniques and a cell-level battery equivalent circuit model to system-level data, accurate SOC and capacity estimation results are achieved. For SOC estimation, the authors proposed a multi-level dual Sigma point Kalman filter (DSPKF) method [2] with its parameters tuned by a genetic algorithm (GA) [3], using either a cell-level or system-level OCV-SOC relationship [1], cell-level equivalent circuit parameters, system-level current and terminal voltage, and a battery equivalent circuit model, for BESS SOC estimation. For capacity (SOH) estimation, system-level current and SOC are inputs of the TLS algorithm [4].

To overcome the limitations of using real-world system-level data, where data granularity, accuracy and quality are limited compared with commonly presented cell-level data, data-cleansing and data-selection techniques are detailed in this paper. The paper first demonstrates the impact of invalid data for SOC estimation accuracy and how to solve this problem, after which, it shows how the quality of data affects SOH estimation accuracy and how to select datasets for improved accuracy.

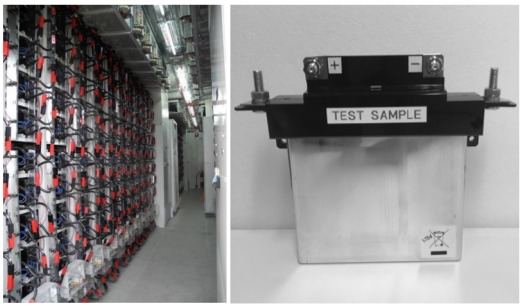


Fig. 1: Photographs of the WESS (left) and single LTO cell (right).

A. Willenhall Energy Storage System

The large-scale battery system studied in this project is the Willenhall Energy Storage System (WESS) operated by the University of Sheffield (UK) which was commissioned in 2016, it has a maximum power of 2MW and a capacity of 1600 Ah (1 MWh). There are 21,120 Toshiba Lithium Titanate Oxide (LTO) cells [5] in this system and each of them has a nominal capacity of 20 Ah. The system is connected to the national electrical grid at 11kV and is used as a research platform for energy trading and providing frequency response services [6]. Fig. 1 shows the photographs of the system and a sample cell. The system is embedded with a Battery Management System (BMS) that provides SOC estimation based on the OCV-SOC relationship and Coulomb counting, which can be used as a reference for KF based SOC estimation and one of the inputs of the TLS algorithm.

B. Battery system SOC estimation using Kalman filter methods

A wide variety of SOC estimation algorithms have been developed in the literature [7]. Among them, KF methods were first proposed for battery cell SOC estimation in [8], [9], by using battery OCV-SOC relationship, terminal voltage and current. The author proposed extended KF and Sigma point Kalman filter [2], [10] successively for battery applications since batteries are non-linear systems. Battery equivalent circuit parameters are variables, so SOC results can be more

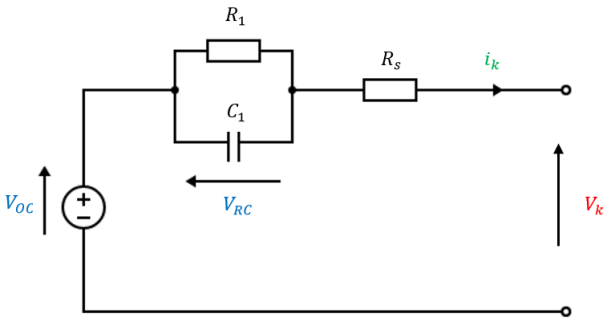


Fig. 2: Equivalent circuit model of WESS [1].

accurate if a second KF is used to estimate these parameters. Therefore, the dual extended Kalman filter (DEKF) and the DSPKF that have such two filters in parallel are recommended. DSPKF as the state-of-art of the KF family is used in [1]. For LTO cells, a 1-RC equivalent circuit model is considered to be efficient and sufficient for SOC estimation according to [11]. The model is shown in Fig. 2 and it is also used for system-level SOC estimation. By understanding the internal structures of a battery system (topology of the cells connected to form the battery), the system-level OCV-SOC relationship and ECM parameters, the main difference from the cell-level estimations, can be calculated. Provided with system-level data, cell-level results, the cell-level model, and equations in [1], a DSPKF SOC estimation that is more accurate than the BMS SOC can be obtained.

C. Battery system capacity estimation using total-least square method

A range of SOH or capacity estimation algorithms have been proposed at the cell-level [12]. Least-square based algorithms, which are relatively simple to implement, can be used to estimate battery capacity, as detailed in [4]. The heart of this algorithm is to use the relationship between the variation of SOC and current integration, which is shown in the equation below:

$$\int_{t1}^{t2} \frac{-\eta I(\tau)}{3600} d\tau = Q(SOC(t_2) - SOC(t_1)) \qquad (1)$$

where η represents the Coulombic efficiency and assumed to be 100% [13], I the charge or discharge current where discharge current is defined to be positive, and Q is the capacity value for calculation.

Dividing the system-level current integration and SOC variation data into a number of segments, using the TLS equations detailed in [4] recursively, battery system capacity can be calculated.

II. Invalid data problem of SOC estimation

The data of the WESS is sometimes invalid, meaning that during these periods the current, voltage and BMS SOC are shown as zeros. This can cause a problem that the KF that is estimating SOC or SOH diverges when the voltage and current values read as zero and continue for sustained period. The reasons for invalid data are as follows:

- A problem of data connection but the battery is still operating.
- The data connection working but the battery is offline (disconnected from the grid and BMS off).
- Both the data connection and the battery are offline.

During the invalid data periods, the EKF and DSPKF algorithms' accuracy is impacted since the invalid data periods start, to be shown in Fig. 3 and Fig. 4. Therefore, making sure the algorithm is able to converge after invalid data is the aim of this work.

A. Methodology

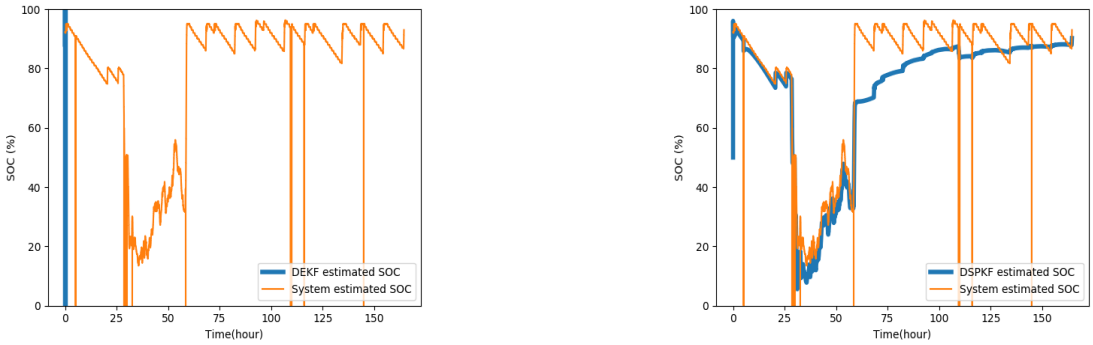
The methods in the literature for invalid data input to a KF mostly concern control and communication, and they often treat the invalid data as a Bernoulli process [14] (a finite or infinite sequence of binary random variables). As it is evident that the data of the WESS does not follow a Bernoulli process, methods should be tried using some empirical knowledge. There are several methods that could solve this problem:

- Use the previous sample point to fill the invalid sample points;
- Pause the KF algorithm, use the last estimation and resume estimating once valid data is received again;
- Limit the estimated SOC values, as the actual SOC values must be between 0% and 100%, to avoid divergence;
- Apply curve fitting using previous sample points to predict the SOC values during the invalid data period.

B. Results

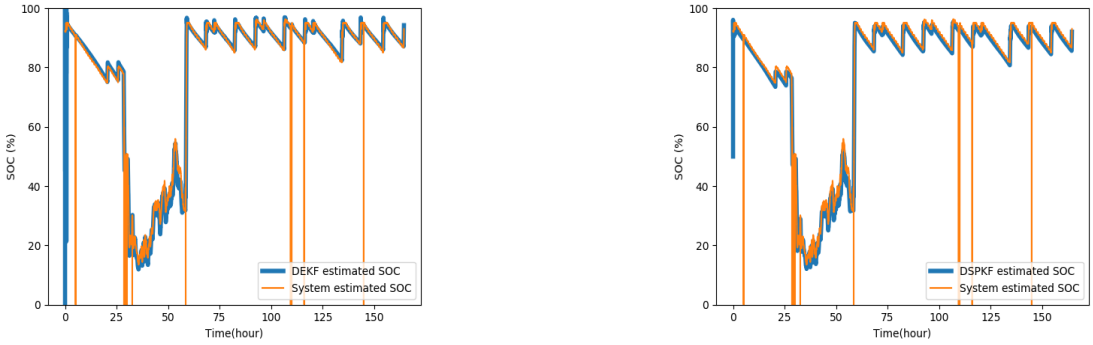
SOC estimation results showing the impact of invalid data and the effectiveness of invalid-data techniques are shown in Fig. 3 and Fig. 4. In Fig. 3 (a), there are several invalid data periods in this profile, but each one was relatively short. The DEKF diverged since the first invalid data period as the estimation started to be minus, and it did not converge again after that as there are no estimates on the figure (which means the estimates of SOC are not within the range of 0% and 100%). In (b), it can been seen that the effects of invalid data are smaller for the DSPKF, but still result in some divergence, and the SOC estimation did not converge after invalid data periods because the weight filter that estimates parameters diverged.

In Fig. 4 (a) and (b), and Fig. 5 (a), the SOC estimation results match well with the BMS SOC, without the effects of invalid data. This is because one of the first two methods is used, together with method three (limiting the SOC). The reason of using one of the first two methods is that they have the same effects to the filter and it can be seen that results in Fig. 4 (a) and Fig. 5 (a) are very similar. If the data starts with invalid data, the SOC estimation should be set as an arbitrary value between 0% and 100% because neither previous estimation or previous data points are available. Using method three (limit SOC range) only is effective but



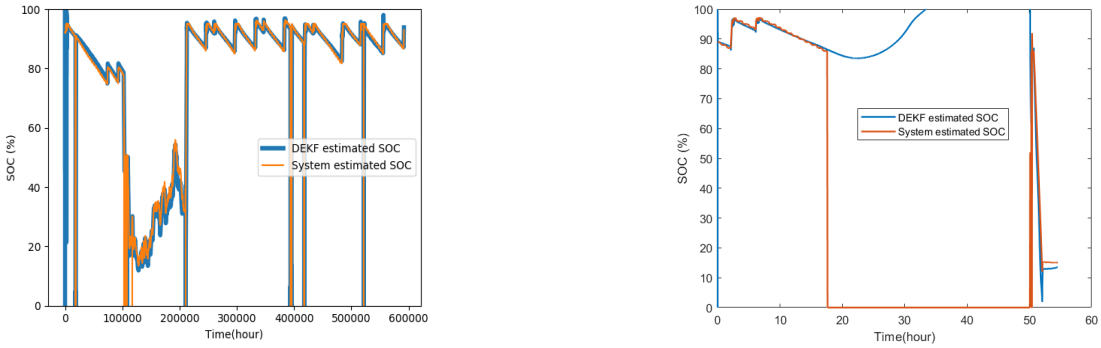
(a) (b)

Fig. 3: SOC estimation results affected by invalid data (a) DEKF SOC and (b) DSPKF SOC.



(a) (b)

Fig. 4: SOC estimation results after using previous data during invalid-data periods (a) DEKF SOC and (b) DSPKF SOC.



(a) (b)

Fig. 5: DEKF invalid-data results with other methods (a) pause the algorithm and (b) curve fitting.

it needs more time to converge after invalid-data periods than the previous methods. This also applies to other circumstances when the calculated SOC is not in this range, which may be due to some very rare erroneous data where there is a time jump after restarting the BESS, the SOC value should be forced back to the reasonable range stated above to avoid further divergence.

The curve fitting method was tried too with SOC limitation using DEKF. The results in Fig. 5 (b) show that the filter still diverges and will exceed the boundaries without SOC limitation. The reason it can converge after the very long time of invalid data is because the SOC is bounded so that any values that larger than 100% are forced to be 100%. The computation time of the algorithm using the curve fitting method is longer than the first two methods due to increased number of calculations. This is exaggerated when there is a very long period of invalid data.

C. Summary

The main problem to solve is that the KF diverges when invalid data happens (both voltage and current values received are 0) if no methods are used, and it may not be able to converge again after the invalid data. It has been shown that Kalman filtering can converge well after the invalid data using the simplest methods. In addition, the length and frequency of the invalid data period do not affect the results of the simple methods. These methods can eliminate the effects of invalid data easily because of the excellent convergence ability of the KF, as long as the previous SOC estimation is forced to be within the reasonable range, i.e., not being negative or larger than 100%.

In conclusion, there are two effective steps to make sure the KF remains converged 1) using the previous data simply when the data is invalid, or maintain the last SOC estimation before invalid data occurs 2) bounding the SOC. The DEKF and DSPKF SOC estimation algorithms are running online (in real-time) for the WESS and have shown robust results against invalid data.

III. Data-selection for online battery system capacity estimation

The TLS algorithm has been used successfully on the system-level capacity estimation as shown in [1]. In the paper, the authors have enumerated the criteria of data that should be chosen in order to maximise capacity estimation accuracy. This data-selection is possible due to the fact that a battery's capacity is not fast-changing and therefore the time between estimates can be large. In this paper, more details, i.e., the essence and methodology of data-selection for online battery system state estimation are shown.

The implementation of "online" capacity estimation algorithm is essential for monitoring the degradation of BESSs during long time operation. For online estimation of the WESS, the data is first obtained from a time-series database (InfluxDB), followed by the TLS capacity estimation algorithm.

A. Methodology

Fig. 6 shows some capacity estimation results using two short datasets [1], where it can be seen that the estimation accuracy differs between them. In (a), after spikes at the beginning during convergence, the results are stable and close to the reference. In (b), the errors are significantly larger and the results fluctuate. This is because the first dataset contains large SOC variations and the current values are mostly constant. Whereas in (b), the dataset is from a frequency response service, which causes small SOC variations and some large and short spikes of current that leads to inaccurate current integration.

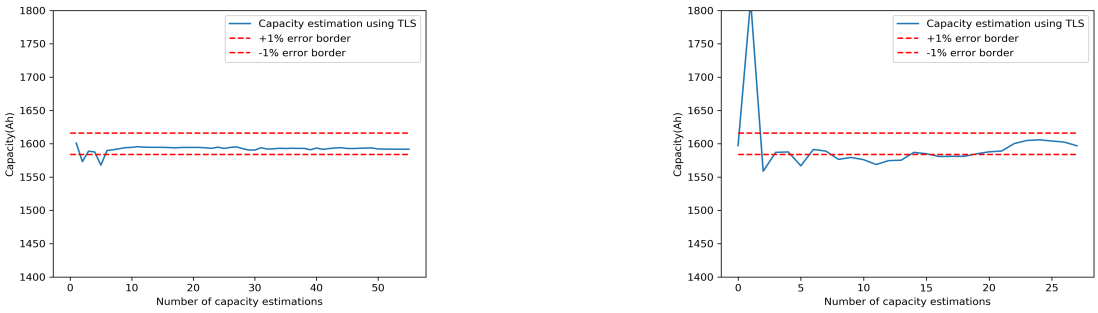
According to the results shown in Fig. 6 and the authors' other experience, a series of criteria of data are shown below for accurate online capacity estimation [1]:

- Significant variations in SOC data are available continuously, as discussed above.
- A large interval size for calculation: as the data is divided into a number of intervals with a size of m, its value should be large enough to make sure there are some SOC variations for every calculation. This value is set to 500 samples for the results shown in Fig. 6.
- Sufficient data: the data points should be enough for the algorithm to converge, since it calculates the capacity recursively.
- No sharp, short spikes of current data: as discussed, to avoid errors in current integration.

An algorithm for data-selection has been developed to select the data that meet the aforementioned criteria. The data (system-level current and SOC) is first divided into chunks representing approximately a week of operation. Next, within each chunk, invalid data is checked for by calculating the change in voltage. When the voltage drops to 0, an invalid data period starts. Likewise, when it returns to within normal operating bounds the invalid period ends. The data chunk is then divided further by deleting the invalid data periods into several even shorter datasets. The length of the first dataset is checked, if it is shorter than the predefined criterion (for the WESS this is 30000 data points, approximately half a day), it is discarded. The algorithm then moves to check the next dataset to ascertain whether the SOC variations are too small, i.e., the standard deviation (STD) is smaller than the predefined value. For similar reasons, to make sure the BESS is operating for services, the STD of current is also checked. After that, the data is checked for any further erroneous parts, for example, the current value being out of bounds. If one of these criteria is not met, the script moves to the next data chunk, until it finds the data to represent this week. The chosen data are then provided to the TLS algorithm introduced in I-C to generate the capacity estimation of this week. Therefore, it is possible that no data is chosen in a week and no capacity estimation results are updated.

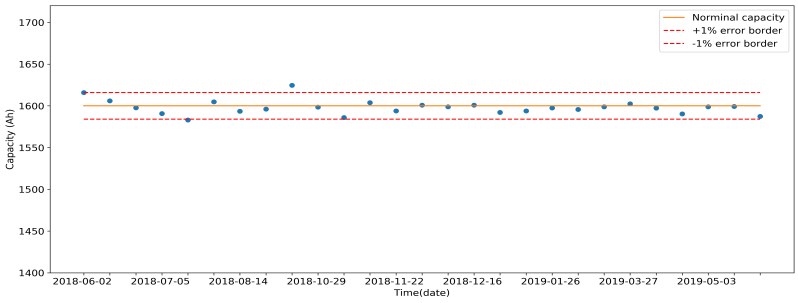
B. Results

By running the algorithm like this, the capacity estimation results are updated every week if there is data that meets the

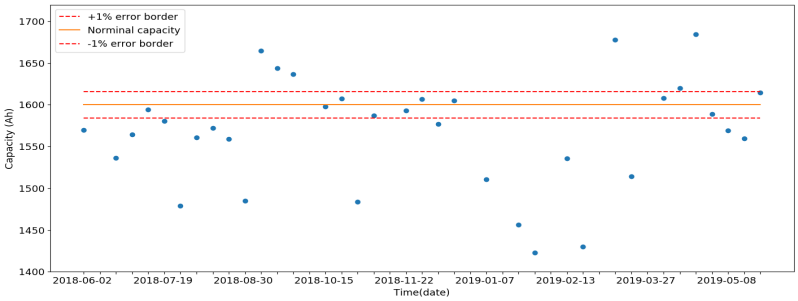


(a) (b)

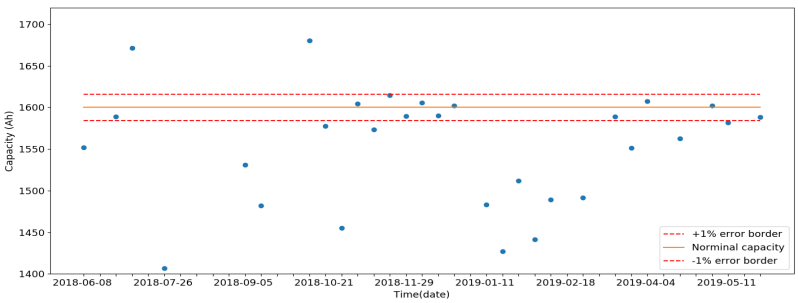
Fig. 6: Capacity estimation results of (a) mixed profile ($\sim$ 9 hours) and (b) dynamic frequency response ($\sim$ 7 hours). The red dotted lines show $\pm 1\%$ error around the 1600 Ah assumed capacity [1].



(a)



(b)



(c)

Fig. 7: Capacity estimation result comparison using a whole year data of 2018 (a) with data-selection techniques, (b) only deal with invalid data and (c) use raw data.

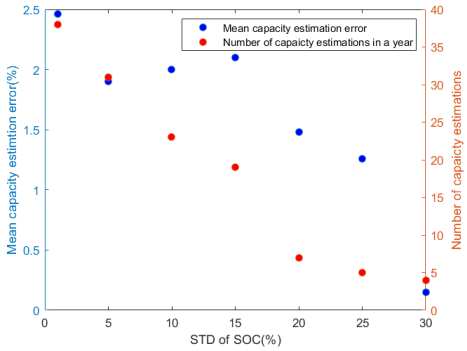


Fig. 8: Effects of SOC variation.

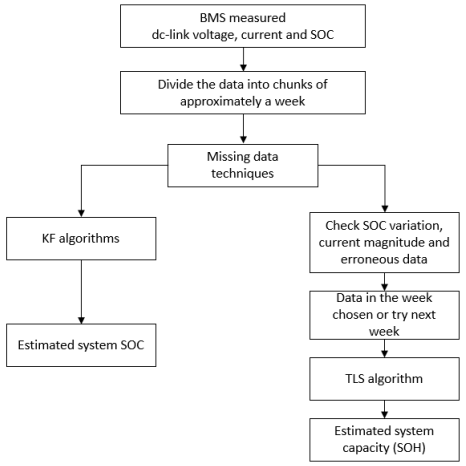


Fig. 9: Diagram of data-selection for BESS SOC and capacity estimation.

criteria. Fig. 7 compares capacity estimation results to show how data-selection improves the accuracy. DSPKF SOC, that has been shown to be more accurate than the BMS SOC, is chosen as the input in these results. Note that the LTO cells in the WESS are still with 100% SOH because of the excellent cycle life of this chemistry. Therefore, the nominal capacity, 1600 Ah is considered as the actual capacity for reference.

It can be seen that if no data-selection is applied at all, only a small fraction of results are within the 1% error borders as shown in Fig. 7(c). Note that extremely large or small results are not shown in the figures. After removing invalid data periods, the results are improved and more are within the error borders and shown in the figure of Fig. 7(b). After applying the data-selection algorithm, the capacity estimation results are mostly within the error borders as shown in Fig. 7(a). To sum up, data-selection techniques ensure the accuracy of TLS algorithms for capacity estimation of the WESS.

Fig. 8 illustrates the effects of SOC variation on one-year's capacity estimation results, using system-level current and the BMS SOC. The mean capacity estimation error is calculated as the quotient of average capacity errors in the year and the actual capacity. In these results, only SOC variation is different, and for comparison, other factors (m and data length) are fixed. It can be seen that generally larger SOC variation improves capacity estimation accuracy, but limits the number of capacity estimation opportunities. To sum up, data-selection techniques ensure the accuracy of TLS algorithms for capacity estimation of the WESS.

IV. Conclusions

This paper presents a method to overcome the invalid data problem in BESS for SOC estimation, using simple techniques to avoid divergence of the KFs. Moreover, it demonstrates how to evaluate the quality of current and SOC for capacity estimation using TLS. The results using the data-selection techniques show significant improvements of accuracy compared with using the data in its raw form. A flow diagram for data-selection for SOC and capacity estimation is shown in Fig. 9: combining state estimation algorithms with data-selection techniques, state estimation of BESSs with improved accuracy using system-level data can be realised.

Acknowledgment

The work within this paper has been supported by Siemens through an ICASE EPSRC grant No. EP/R512175/1.

References

[1] Z.Wang, D. T. Gladwin, M. J. Smith, and S. Haass, "Practical state estimation using kalman filter methods for large-scale battery systems," *Applied Energy*, vol. 294, 2021.

[2] G. L. Plett, "Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs part 1: Introduction and state estimation," *J. Power Sources*, vol. 161, no. 2, pp. 1356–1368, 2006.

[3] F.-A. Fortin, F. M. D. Rainville, M. A. Gardner, M. Parizeau, and C. Gagné, "Deap: Evolutionary algorithms made easy," *J. Mach.Learn. Res*, vol. 13, no. 1, pp. 2171–2175, January 2012.

[4] G. L. Plett, "Recursive approximate weighted total least squares estimation of battery cell total capacity," *J. Power Sources*, vol. 196, no. 4, pp. 2319–2331, 2011.

[5] Toshiba, "SCiB cells," https://www.scib.jp/en/product/cell.htm, 2019.

[6] NationalgridESO. (2021) Frequency response services. https://https://www.nationalgrideso.com/industry-information/balancing-services/frequency-response-services.

[7] M. A. Hannan, M. S. H. Lipu, A. Hussain, and A. Mohamed, "A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations," *Renew. Sustain.Energy Rev.*, vol. 78, pp. 834–854, October 2017.

[8] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs - part 1. modeling and identification," *J. Power Sources*, vol. 134, no. 2, pp. 262–276, August 2004.

[9] ——, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs, part 3: State and parameter estimation," *J. Power Sources*, vol. 134, no. 2, pp. 277–292, August 2004.

[10] ——, "Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs part 2: Simultaneous state and parameter estimation," *J. Power Sources*, vol. 161, no. 2, pp. 1369–1384, 2006.

[11] S. Nejad, D. T. Gladwin, M. P. Foster, and D. A. Stone, "Parameterisation and online states estimation of high-energy lithium-titanate cells," *IECON*, November 2017.

[12] R. Xiong, L. Li, and J. Tian, "Towards a smarter battery management system: A critical review on battery state of health monitoring methods," *J. Power Sources*, vol. 405, pp. 18–29, November 2018.

[13] F. Yang, D. Wang, Y. Zhao, K. Tsui, and S. J. Bae, "A study of the relationship between coulombic efficiency and capacity degradation of commercial lithium-ion batteries," *Energy*, vol. 145, pp. 486– 495, 2018.

[14] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, September 2004.