Article:
Wang, Z., Gladwin, D.T., Smith, M.J. et al. (2021) Practical state estimation using Kalman filter methods for large-scale battery systems. Applied Energy, 294. 117022. ISSN: 0306-2619

https://doi.org/10.1016/j.apenergy.2021.117022

Practical state estimation using Kalman filter methods for large-scale battery systems

Zhuo Wang[a], Daniel T. Gladwin[a,*], Matthew J. Smith[a], Stefan Haass[b]

[a] *Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom*
[b] *Siemens plc, RC-GB SI DS FG, 2 Koppers Way, Hebburn NE31 2EZ, United Kingdom*

Abstract

The system states of battery energy storage systems (BESSs) such as state of charge (SOC) and state of health (SOH) are essential for the functions of the system, such as frequency support services and energy trading. However, the complexity of a large-scale battery system makes the estimations more difficult than at the cell-level. This is further compounded by real-world limitations on system monitoring data granularity, accuracy and quality. In this paper it is shown how cell-level state estimation techniques can be utilised on large-scale BESSs using experimental data from a 2MW, 1MWh BESS. The results show how a Dual Sigma point Kalman Filter (DSPKF) SOC estimation provides more accurate results compared to the commercial BESS battery management system SOC. It is shown how the DSPKF parameters can be tuned by a genetic algorithm to simplify selection and generalise the approach for different BESSs. Furthermore, it shows how this method of SOC estimation can be combined with a total least-squares (TLS) method for capacity estimation to less than 1% error. Online system state estimation is demonstrated using both designed tests and real-world operational profiles where the BESS has provided contracted frequency response services to the national electricity grid in the UK.

Keywords: Battery system, State of charge, State of health, Kalman filter,

*Corresponding author
Email addresses: zhuowang@sheffield.ac.uk (Zhuo Wang), d.gladwin@sheffield.ac.uk (Daniel T. Gladwin), matt.j.smith@sheffield.ac.uk (Matthew J. Smith), stefan.haass@siemens.com (Stefan Haass)

Nomenclature

ABS Absolute value

ANN Artificial neural networks

$BESS$ Battery energy storage system

BMS Battery management system

$DEKF$ Dual extended Kalman filter

$DSPKF$ Dual Sigma point Kalman filter

GA Genetic algorithm

KF Kalman filter

LTO Lithium titanate

OCV Open-circuit-voltage

$RMSE$ Root mean square error

SOC State of charge

SOH State of health

$SPKF$ Sigma point Kalman filter

TLS Total least squares

$WESS$ Willenhall Energy Storage System

1. Introduction

Global warming and pollution caused by burning fossil fuels is widely recognised as an imminent threat to the planet[1]. Renewable energy sources like wind and solar power can help to mitigate these problems, but their unpredictability and variability of generation capacity can cause instabilities in voltage and frequency of electricity networks[2]. The amount of power generated from these renewable sources depends on natural conditions, where generation exceeds demand, reliable methods are needed to store this excess energy. Conversely, where demand exceeds supply, energy storage may be used to provide support and stability to the grid whilst additional generation capacity becomes available.

Large-scale battery energy storage systems (BESSs) have recently emerged as a solution to provide a variety of grid support services [3], thereby helping to mitigate the instability problems from renewable energy resources noted above. A range of UK national grid frequency response services such as firm frequency response and fast reserve can be achieved by BESSs [4], they can also be used to achieve price arbitrage and balancing services. The recent advances in battery chemistry technologies have improved the performance of BESS in terms of higher volumetric energy capacities, better round-trip efficiencies, and longer lifetime. To make effective use of these advances, to successfully provide grid support and maximise the return on investment for battery owners, advanced battery management systems (BMS) need to be developed.

State of charge (SOC) and state of health (SOH) are the two essential indicators that need to be estimated by the BMS. Capacity, that represents the maximum electrical charge that a battery can store presently, is directly related to these two indicators. SOH is the quotient of the actual capacity and the nominal capacity, whereas, SOC is the percentage of charge held by the battery presently with respect to the actual capacity. The capacity of a battery reduces over time and its rate of degradation is predominantly dependant upon the type of usage. The accuracies of SOC [5] and SOH [6] estimations at the cell-level

have improved significantly over recent years from numerous researchers. However, for large-scale battery systems, the accurate estimation of SOC and SOH is a relatively new topic.

The accurate estimation of a BESS's capacity and SOC are critical for its operation. Batteries have limited lifespans; when the criterion of the end of life, often around 70% or 80% of nominal capacity, is reached, the battery will no longer serve the need of the application. Moreover, as the efficiency of the battery decreases, there is a higher risk of permanent failure of cells within the battery system. As providing services reduces the SOH, the service must be financially beneficial when taking into consideration the loss of capacity incurred. Accurate SOH prediction allows for an estimation of the cost in terms of battery degradation of a service to be compared against the profit earned by performing the service. In this way, it is possible to optimise the control and availability of the services provided, to generate maximum profit whilst causing minimum degradation of the BESS. It is also important to be able to predict the capability of the system to store energy going forward, thus being able to avoid services that require more energy than the BESS can provide. The necessities for accurate SOC estimation are more straightforward: the owner of a BESS needs to know how much energy is currently stored to provide grid services and actively manage the SOC to remain within the service requirements.

There are already a wide variety of methods to estimate battery capacity and SOC, but most of them focus at the cell-level (section 2). Plett. et al. [7] proposed a Bar-Delta filtering method for estimating the battery pack states, not only estimating the pack-average states but also the differences between the cell and pack-average states. In [8], an online estimation technology based on KF methods for SOC of all cells in the pack was proposed. Authors in [9] achieved accurate SOC estimation of a battery pack using an adaptive extended Kalman filter. Although the above methods could work well for battery packs, they would be much slower to be implemented for large-scale battery systems due to the significantly higher number of cells. Recently, a digital twin method [10] for online SOC and SOH estimation was proposed. A series of studies on a sample of

Spanish photovoltaic household-prosumers [11] showed factors that could affect a battery system's lifetime [12] and predicted a battery bank's lifetime based on fixed battery degradation parameters, under fluctuating loads [13]. To conclude, battery systems that previous studies use to demonstrate their methodologies contain a very small number of cells relative to large-scale grid connected BESS. A large-scale BESS system will be made up of storage units that can contain in excess of 20k-100k cells each and with it come challenges in measurement, data granularity, accuracy and data quality. The state estimation of an entire large-scale BESS for increased accuracy, beyond commonly used methods, has not been presented in the literature.

This paper first demonstrates the implementation of a Dual Sigma point Kalman (DSPKF) on a BESS for SOC estimation, and exploits a genetic algorithm (GA) to solve the parameter selection problem. For online system capacity estimation, a total least-squares (TLS) solution is experimentally demonstrated for a large-scale BESS that will highlight the need for data selection and cleansing. The TLS solution requires an SOC estimation input, to demonstrate the accuracy of the DSPKF alogirthm this is used and compared against the BMS provided SOC. Finally, a method to improve SOC estimation and capacity accuracy as the BESS degrades is proposed. System-level OCV-SOC and degradation experimental results are also presented to validate the system-level state estimation results. Fig. 1 provides an overview of how the various methods interact, with the numbers in each box indicating the relevant sections in this paper. All data from the BESS is obtained at the highest level in the system as compared to interfacing at lower levels that would produce faster sampled and consistent data. This is representative of asset owner access to data in a real-world system and demonstrates the potential for the methodologies presented in this paper to process data both locally and remotely, for example in cloud based systems.

2. Methods for SOC and SOH Estimation

Commonly cited methods for SOC and SOH are listed below:

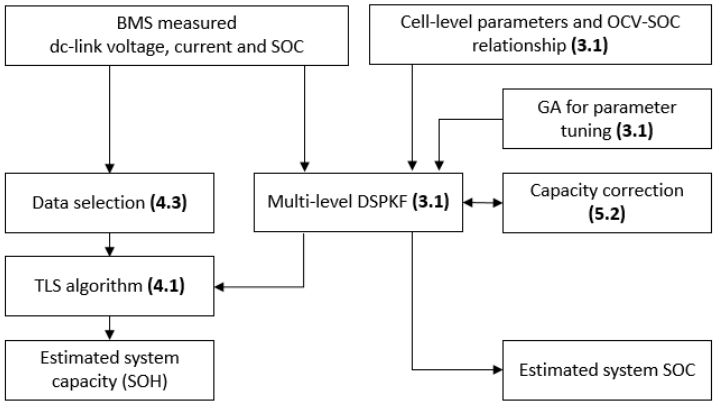


Fig. 1: Battery system state estimation flowchart

i Coulomb counting;

ii Kalman filter algorithms;

iii Machine learning algorithms;

Coulomb counting is the most often used method in industry and commercial applications to estimate SOC. The equation of calculating SOC is shown in Eq. (1)[14]. The change in SOC is calculated by accumulating the charge transferred in or out of the battery, therefore, the initial SOC must be available to estimate SOC in this method.

$$SOC(t_0 + \tau) = SOC(t_0) + \frac{1}{C_{rated} SOH} \int_{t_0}^{t_0+\tau} -I dt \tag{1}$$

where $SOC(t_0)$ is the initial SOC, C_{rated} the rated(nominal) capacity, τ the time duration of charge or discharge, and I is the input or output current. Note that in this paper, discharge current is defined as positive, which is in line with convention.

Coulomb counting may be suitable for capacity estimation only when a full discharge is available (or almost full discharge, to be proposed later) and when accurate current sensors are available. This method cannot be utilised when a BESS is in operation for grid services and therefore requires the BESS to

go offline to carry out a full discharge. Similar to Eq. (1) , the calculation of capacity is shown in Eq. (2) by integrating the discharge current.

$$C = \int_{t_{start}}^{t_{end}} I_D dt \tag{2}$$

where t_{start} is the start time of the discharge process and t_{end} is the end time. I_D is the discharge current.

The advantage of Coulomb counting is the simplicity and that it is a direct method, with the SOC and capacity can be estimated requiring only the measurement of the current [15]. However, it can be very inaccurate. There are losses during charging and discharging, and these, in addition to self-discharging, result in errors. The measurement of current is another problem since the current sensors can be affected by offset errors or noisy measurements, and these combined errors accumulate into increasingly large errors as time passes. The estimated SOC will diverge from the actual SOC, although a reset mechanism using the open-circuit voltage and SOC relationship (OCV-SOC relationship), discussed in more detail in section 3 can mitigate this but only under ideal conditions not necessarily witnessed by an online operational BESS.

Kalman filter (KF) methods can provide very accurate indirect estimations of SOC. They compute a weighted average of the measured value and the predicted value by utilising a set of recursive equations to minimise the noise values [16]. The weighted average (Kalman gain) is calculated by placing heavier weights on more likely values according to the error covariances. Estimates of both the state and the error covariance matrices to be minimised are the heart of the solution in KF methods. Prediction and correction are the two main steps in KF: in the prediction step, the state of the system is estimated using the previous measurement, while in the correction step, the estimated state is updated with the measurement.

Battery systems are nonlinear so a standard (non-extended) Kalman filter cannot be used. The Extended Kalman filter (EKF) was first proposed [16], but the Sigma-point Kalman filter (SPKF) [17] is more advanced than EKF, and is seen as the state-of-art of KF algorithms. SPKF computes the covariance

matrix by using the results of a number of function evaluations, which decrease the error of linearisation significantly. Moreover, it can achieve higher accuracy with similar computational complexity to EKF. There are various sigma-point methods, and among them the Central Difference Kalman Filter (CDKF) is chosen in this work because it is simpler to be implemented and has higher theoretical accuracy [18].

Dual Sigma point Kalman filter (DSPKF) [19] and dual extended Kalman filter (DEKF) [20] can realise the estimation of both state and parameter values with two separate filters. The two filters in these methods are called state filter and weight filter respectively, and the weight filter is designed for equivalent circuit parameter estimation since the parameters only vary slowly with time for a battery. Therefore, DSPKF is more accurate than a single SPKF that has fixed parameters. The two filters run in parallel, they adapt the parameters and the state respectively with some information exchange.

There is a range of studies that use KF methods to estimate SOC and capacity. In [16], a KF was first proposed for battery applications, the inputs include the current, voltage and temperature experienced by the cell, and the output is the SOC. The SOC is first predicted using a battery model then, the open-circuit voltage can be formulated according to the OCV-SOC relationship. Thereafter, the OCV is used to calculate the terminal voltage, which is compared with the measured voltage to correct the prediction. Zou et al.[21] proposed a novel combined SOC and SOH estimator, and in this work, SOC was estimated in real-time using a second-order EKF (two state variables), and SOH was updated offline using a fourth-order EKF. Authors in [22] proposed a multi-scale DEKF algorithm for lithium-ion batteries to significantly reduce the computational burden, based on that the parameters estimated in the weight filter do not change fast.

KF methods are not complex and can be implemented with a systematic approach offering high accuracy and robustness against poor initialisation. The disadvantages are that they are sensitive to modelling accuracy; the battery operation environment should be in the zero-mean noise condition (Gaussian)[23].

Data-driven methodology has drawn significant attention recently, thanks to the rapid development of machine learning. A complicated but accurate model of a system can be trained first, using sufficient data, after which the model is used to predict the required values, such as SOC and SOH in battery systems. These methods can be even more accurate than the Kalman filtering, including artificial neural network (ANN), support vector machine (SVM) etc. [24], but they require a significant amount of data and demand higher computation. They are frequently applied for the prediction of SOH and remaining useful life (RUL)[25].

Among various SOH only algorithms [26], total least-square based methods can be relatively simply implemented, without the sacrifice of accuracy, if the data quality is guaranteed. In [27], the relationship between current integration and SOC variation was used as the foundation of total least-square based methods for capacity estimation. These algorithms, which include weighted total least squares (WTLS), total least squares (TLS) and approximate weighted total least squares (AWTLS), attempt to find an estimated capacity that minimises the sum of squared errors, which is done recursively. TLS is used in this work for system-level capacity estimation and the details of applying this method are in section 4.

3. DSPKF implementation on large battery system

3.1. The model and equations

In this chapter cell-level DSPKF techniques are going to be demonstrated on a large scale BESS called the Willenhall Energy Storage System (WESS). WESS is the largest research based grid-connected lithium titanate energy storage facility (11 kV) in the UK [4], which was launched in 2016. Lithium titanate (LTO) batteries are safer, offer high charge/discharge rates, low-temperature operation and significantly longer lifespans [28] by comparison with conventional lithium-ion batteries. The disadvantages of these cells are their higher cost and that they operate at a lower voltage, however this can be overcome in BESSs by connecting more cells in series.

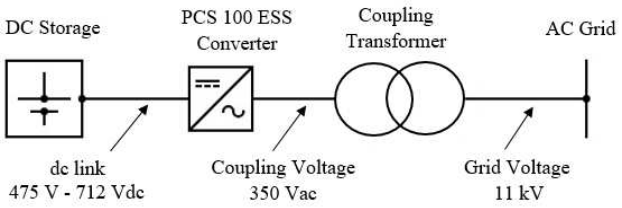


Fig. 2: System diagram of WESS.

The system consists of 21,120 Toshiba LTO cells and the highest power and energy throughput are 2 MW and 1 MWh respectively. The nominal capacity of a single cell is 20 Ah and for the system is 1600 Ah. The system consists of 40 racks in parallel, 22 modules in series in every rack, and 24 cells in each module in a 2P12S configuration. The modules in each rack communicate over CAN bus to a rack management unit (RMU) that collects the voltages of all the parallel connected cells (12 voltage measurements). The RMU also measures the total current in/out of each rack, which is then communicated back over CAN bus along with voltage information (cell Min/Max and rack voltage) to a system BMS. The system BMS measures the overall dc-link voltage and current that is connected to the inverter. The BMS reports over Modbus TCP/IP various system parameters, those important for this paper are, dc-link voltage, current and SOC. An air-conditioning system maintains the temperature of the cells to a narrow range of between 22-30 °C under normal operation and therefore the effects of extended temperatures are not considered in this paper. A diagram of the system is shown in Fig. 2.

The BMS reported SOC of WESS is an estimate using a Coulomb-counting method and uses the OCV-SOC relationship for correction when the battery is at a predefined condition during voltage relaxation. As previously discussed, the SOC therefore suffers from error accumulation over periods of sustained charge/discharge. The DSPKF method, introduced earlier, has achieved high accuracy for SOC estimation at the cell-level in the literature. In this section it is shown how this method can be applied to a large-scale BESS such as WESS.

Table 1 shows the equations for a multi-scale DSPKF implementation [19,

22]. They describe the whole procedure of initialisation, time update (prediction) and measurement update (correction) for both state and weight filters. k is the sample rate of the system data, and m is the macro scale, which is a quantity of samples. The parameters in the first order equivalent circuit model and the battery capacity are slowly time-varying so using a macro scale to estimate them is deemed efficient [29]. The SOC is estimated every sample while the parameters including capacity are estimated each time there are m new data samples.

In Table 1, forms of X are for the states estimation (SOC), forms of θ battery equivalent circuit parameters, forms of L the Kalman gains, forms of P the error covariance and noise covariance matrices, and E represents the expectation. u_k and y_k are the measured current and voltage respectively. w_k and v_k are the process and observation noises. For the calculations of $\alpha_i^{(m)}$, $\alpha_i^{(c)}$ and lower triangular matrix see [19]. The state vector is augmented to include the noise effects so that the new vector length is p. $f(\bullet)$ and $h(\bullet)$ are the non-linear state transition and observation models' functions, respectively. They are derived from the first-order equivalent circuit model [30] of a battery cell for the battery system, as shown in Fig. 3.

Table 1: Multi-scale DSPKF implementation

Definitions:

$$x_k^a = [{x_k}^T, {\omega_k}^T, {v_k}^T]^T, \qquad X_k^a = [(X_k^x)^T, (X_k^\omega)^T, (X_k^v)^T]^T$$

$$p = 2 \times dim(x_k^a)$$

Initialisation: for k=0, set

$$\hat{\theta}_0^+ = E[\theta_0], \qquad P_{\tilde{\theta},0}^+ = E[(\theta_0 - \hat{\theta}_0^+)(\theta_0 - \hat{\theta}_0^+)^T]$$

$$\hat{x}_0^+ = E[x_0], \qquad \hat{x}_0^{a,+} = E[x_0^a] = [(\hat{x}_0^+)^T, \bar{\omega}, \bar{v}]^T$$

$$P_{x,0}^+ = E[(x - \hat{x}_0^+)(x - \hat{x}_0^+)^T]$$

$$P_{x,0}^{a+} = E[(x_0^a - \hat{x}_0^{a,+})(x_0^a - \hat{x}_0^{a,+})^T] = diag(P_{x,0}^+, P_\omega, P_v)$$

State filter equations, for k=1,2,... compute:

Time-update equations for state filter

$$X_{k-1}^{a,+} = \{\hat{x}_{k-1}^{a,+}, \hat{x}_{k-1}^{a,+} + \sqrt{P_{\tilde{\mathbf{x}},k-1}^{a,+}}, \hat{x}_{k-1}^{a,+} - \sqrt{P_{\tilde{\mathbf{x}},k-1}^{a,+}}\}$$

$$X_{k,i}^{x,-} = f(X_{k-1,i}^{x,+}, u_{k-1}, X_{k-1,i}^{\omega,+}, \hat{\theta}_k^-, k-1) \qquad \hat{x}_k^- = \sum_{i=0}^{p} \alpha_i^{(m)} X_{k,i}^{x,-}$$

$$P_{\tilde{\mathbf{x}},k}^- = \sum_{i=0}^{p} \alpha_i^{(c)} (X_{k,i}^{x,-} - \hat{x}_k^-)(X_{k,i}^{x,-} - \hat{x}_k^-)^T$$

Output estimate, state filter

$$Y_{k,i} = h(X_{k,i}^{x,-}, u_k, X_{k-1,i}^{v,+}, \hat{\theta}_k^-, k) \qquad \hat{y}_k = \sum_{i=0}^{p} \alpha_i^{(m)} Y_{k,i}$$

State filter gain matrix

$$P_{\tilde{\mathbf{y}},k} = \sum_{i=0}^{p} \alpha_i^{(c)} (Y_{k,i} - \hat{y}_k)(Y_{k,i} - \hat{y}_k)^T$$

$$P_{\tilde{\mathbf{x}}\tilde{\mathbf{y}},k}^- = \sum_{i=0}^{p} \alpha_i^{(c)} (X_{k,i}^{x,-} - \hat{x}_k)(Y_{k,i} - \hat{y}_k)^T$$

$$L_k^x = P_{\tilde{\mathbf{x}}\tilde{\mathbf{y}},k}^- P_{\tilde{\mathbf{y}},k}^{-1}$$

Measurement-update equations for state filter

$$\hat{x}_k^+ = \hat{x}_k^- + L_k^x(y_k - \hat{y}_k) \qquad P_{\tilde{\mathbf{x}},k}^+ = P_{\tilde{\mathbf{x}},k}^- - L_k^x P_{\tilde{\mathbf{y}},k}(L_k^x)^T$$

Weight filter equations,for $k \mod m = 0$ **compute:**

Time-update equations for weight filter

$$\hat{\theta}_k^- = \hat{\theta}_{k-1}^+ \qquad P_{\tilde{\theta},k}^- = P_{\tilde{\theta},k-1}^+ + P_r$$

Output estimate, weight filter

$$W_k = \{\hat{\theta}_k^-, \hat{\theta}_k^- + \sqrt{P_{\tilde{\theta},k}^-}, \hat{\theta}_k^- - \sqrt{P_{\tilde{\theta},k}^-}\}$$

$$D_{k,i} = h(f(\hat{x}_{k-1}^+, u_{k-1}, \bar{\omega}_{k-1}, W_{k,i}, k-1), u_k, \bar{v}_k, W_{k,i}, k)$$

$$\hat{d}_k = \sum_{i=0}^{p} \alpha_i^{(m)} D_{k,i}$$

Parameter filter gain matrix

$$P_{\tilde{\mathbf{d}},k} = \sum_{i=0}^{p} \alpha_i^{(c)} (D_{k,i} - \hat{d}_k)(D_{k,i} - \hat{d}_k)^T$$

$$P_{\tilde{\theta}\tilde{\mathbf{d}},k}^- = \sum_{i=0}^{p} \alpha_i^{(c)} (W_{k,i} - \hat{\theta}_k)(D_{k,i} - \hat{d}_k)^T$$

$$L_k^\theta = P_{\tilde{\theta}\tilde{\mathbf{d}},k}^- P_{\tilde{\mathbf{d}},k}^{-1}$$

Measurement-update equations for weight filter

$$\hat{\theta}_k^+ = \hat{\theta}_k^- + L_k^\theta(y_k - \hat{d}_k) \qquad P_{\tilde{\theta},k}^+ = P_{\tilde{\theta},k}^- - L_k^\theta P_{\tilde{\mathbf{d}},k}(L_k^\theta)^T$$

There could be n RC branches, and the model accuracy is higher with more branches, but it increases calculation burdens. Both models with 1-RC pair and 2-RC pairs have been tried but the SOC results were very similar, because of the dominant diffusional impedance effects in LTO cells [31]. Thus, the model with only one RC branch is chosen for LTO batteries in the paper, which is

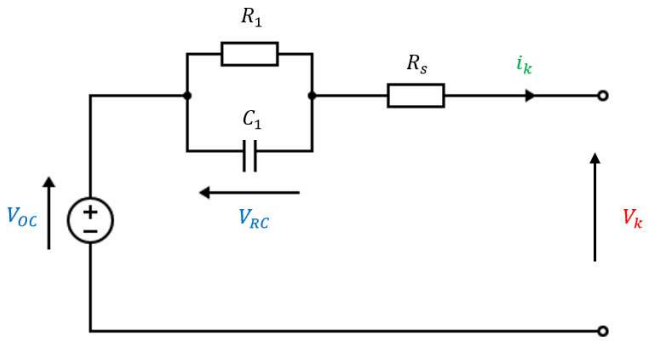


Fig. 3: Equivalent first-order circuit model of WESS.

considered as having sufficient accuracy at low computational complexity.

$$f(\bullet) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} SOC_{K+1} \\ V_{RC_{k+1}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{-\Delta t}{\tau_1}} \end{bmatrix} \begin{bmatrix} SOC_K \\ V_{RC_k} \end{bmatrix} + \begin{bmatrix} -\frac{\eta \Delta t}{Q_{actual}} & 0 \\ 0 & R_1(1 - e^{\frac{-\Delta t}{\tau_1}}) \end{bmatrix} I_k \tag{3}$$

$$\theta_k = [R_s, R_1, \tau_1]^T \tag{4}$$

$$h(\bullet) = V_k = V_{OC}(SOC_k) - V_{RC1_K} - I_k R_s \tag{5}$$

Eq. (3) to Eq. (5) [30] are used for the DSPKF implementation based on the chosen model. η is the battery Coulombic efficiency and Δt is the sample period of data. In the RC branch, the changes in C_1 are suitable indicators of the changes in SOH [32], R_1 the self-discharge resistance, and $\tau_1 = C_1 R_1$ is the time constant of the RC branch. The last equivalent circuit parameter R_s represents the resistance of the battery's terminals and inter-cell connections. Q_{actual} is the actual capacity in Coulombs and V_{OC} (SOC) is the OCV-SOC relationship which is obtained by experiments: the results shown in this work use the scaled OCV-SOC experimentally measured of a single LTO cell, although the system-level OCV-SOC relationship was also experimentally captured as

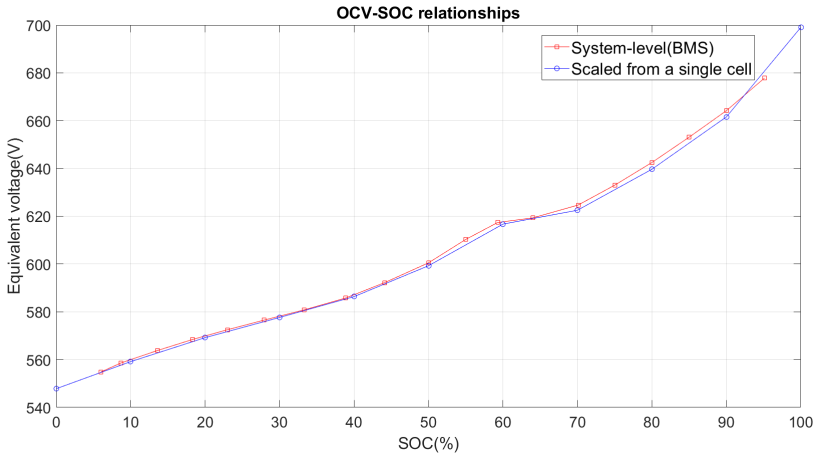


Fig. 4: OCV-SOC relationships: scaled from cell-level vs system-level.

Table 2: GA settings

Parameter	Value
Number of generations	10
No. individuals in a population	200
Selection operator	NSGAII
Independent probability for each attibute to be mutated	0.05
Crossover probability	0.5
Mutating probability	0.1
Fitness function 1	ABS(BMS SOC - DSPKF estimated SOC)
Fitness function 2	ABS(scaled Rs - DSPKF estimated Rs)

shown in Fig. 4. The state vector (to be estimated) contains the SOC and V_{RC1}. The value of Δt depends on the sample rate of the system which for WESS is nominally 1Hz but in reality this varies.

The input parameters of the DSPKF algorithm are sampled current and voltage data, the OCV-SOC relationship, battery Coulombic efficiency, measured capacity of the battery system, initial state and estimated equivalent circuit parameter values. The outputs of the algorithm are the system SOC and updated estimates of the equivalent circuit parameters. The initial equivalent parameter values are selected as scaled values from cell-level electrochemical impedance spectroscopy experimental results. The topology of the cells connected to form the battery is known, it is therefore possible to calculate scaled values from a simple circuit model.

The error covariance matrices for process noise and sensor noise are an important selection for successful implementation of a DSPKF [16]. Genetic Algorithms (GA) are excellent for solving searching and optimisation problems [33] and are viewed as a "universal optimizer". The simplicity and ease of implementation allow, with careful selection of parameters, a good balance of exploration and exploitation of the search space. Whilst other heuristic techniques could be applied to this problem, finding the most efficient heuristic is out of scope of this paper as the optimisation is only ran once. Thus, in this work the DSPKF parameters are automatically tuned by a GA using the Distributed Evolutionary Algorithm in Python (DEAP) library. The GA settings are shown in Table 2. Note that the values in the table are not unique and were discovered through trial and error using the experience of the authors and tracking progress through generations. To make sure a significant proportion of the search space is explored, the population should be large whilst consideration is taken for the computation time. The number of generations and the possibilities do not affect the efficiency of GA significantly for this application. The parameters include elements of process noise matrices in both state and weight filters, sensor noise, elements in error covariance matrices of both filters. The GA is multi-objective with two fitness functions defined and the non-dominated sorting genetic algorithm (NSGA-II)[34] is applied to select the Pareto front as candidates of the optimums. The first fitness function is designed to minimise the mean average error (MAE) between the DPKF SOC estimation and BMS SOC data known to be of acceptable accuracy (cycles that are known by experience of using the BESS not to cause large SOC discrepancies). The second fitness function is defined to minimise the error of the DPKF estimated R_s equivalent circuit parameter against the calculated value. The authors investigated additional fitness functions to include other equivalent circuit parameters but found no significant improvements beyond optimising for R_s.

Applying $f(\bullet)$ and $h(\bullet)$ and equations in Table 1 recursively, with the system parameters and KF parameters stated above, DSPKF can converge to provide an estimated battery SOC.

3.2. Results of DSPKF SOC estimation for WESS

Fig. 5 shows the SOC estimation results of constant cycling, a mixed profile and grid frequency response service. The DSPKF and BMS results are a good match with the RMSE calculated to be smaller than 1.5%. In (a) and (b), the DSPKF and BMS estimated SOC are very close because the two datasets are with almost constant current. In (c) the differences are larger and this is due to the BESS delivering Dynamic Frequency Response (fluctuating load), with various and fast-changing current values, so the BMS fails to provide accurate estimations (demonstrated in section 6). The SOH of the system will affect the SOC estimation accuracy significantly (Eq. (1)). For the shown SOC results, η is assumed to be 100% [35], and the SOH is set to 100% since WESS's SOH is currently still very high and the degradation too small to measure (detailed in section 4).

4. Capacity estimation on large system

4.1. Total least-squares algorithm

Battery capacity can be estimated by total least-square based methods [27] by using the relationship between the variation of SOC and current integration, as shown below:

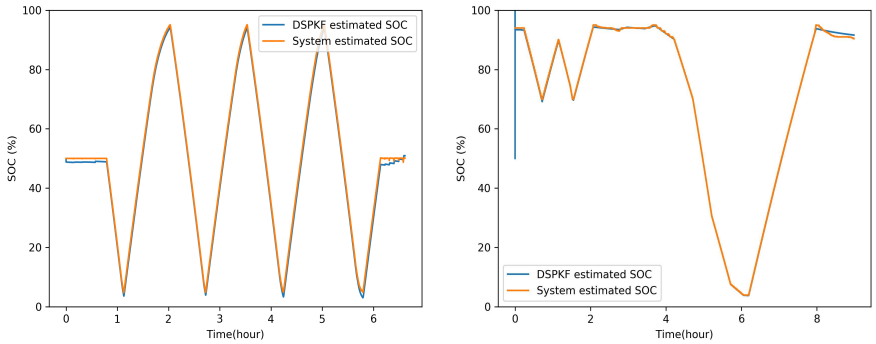
$$\int_{t1}^{t2} \frac{-\eta I(\tau)}{3600} d\tau = Q(SOC(t_2) - SOC(t_1)) \tag{6}$$

where η is again the Coulombic efficiency and assumed to be 100%, I the charge or discharge current and Q is the capacity value that needs to be calculated. This equation is based on Eq. (2), the only difference is that it refers to the condition when a full discharge is not available.

For the simplicity of calculation:

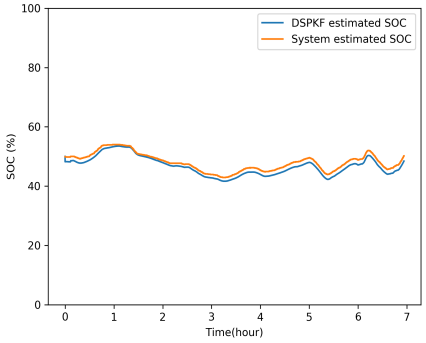
$$y = \int_{t1}^{t2} \frac{-\eta I(\tau)}{3600} d\tau \quad and \quad x = SOC(t_2) - SOC(t_1) \tag{7}$$

The total least squares (TLS) method is used in this work, which assumes errors in both y and x data, but the error variances are proportional. Therefore, $\sigma_{y_n}^2 = k^2 \sigma_{x_n}^2$. The equations below show the iterative calculations of TLS.



(a)

(b)



(c)

Fig. 5: SOC estimation results of (a) constant cycling and (b) mixed profile and (c) dynamic frequency response.

The data is divided into n segments to do the recursive calculations, $\sigma^2_{y_n}$ the error variance in y of every segment and $\sigma^2_{x_n}$ is the error variance in x of every interval. They are guesses of the errors on both y and x respectively. The equations below show the iterative calculations of the method [27].

$$\begin{aligned} c_{1,n} &= c_{1,n-1} + \frac{x_n^2}{\sigma^2_{y_n}}; \\ c_{2,n} &= c_{2,n-1} + \frac{x_n y_n}{\sigma^2_{y_n}}; \\ c_{3,n} &= c_{3,n-1} + \frac{y_n^2}{\sigma^2_{y_n}}; \end{aligned} \tag{8}$$

$$\hat{Q_n} = \frac{-c_{1,n} + k^2 c_{3,n} + \sqrt{(c_{1,n} - k^2 c_{3,n})^2 + 4k^2 c_{2,n}^2}}{2k^2 c_{2,n}}; \tag{9}$$

where $c_{1,n}$, $c_{2,n}$ and $c_{3,n}$ are quantities to help calculations.

4.2. Capacity estimation results using BMS SOC

In this section several datasets of measured current and BMS SOC from WESS are used to estimate the system capacity and compare it with offline experimentally measured capacity using Coulomb counting.

4.2.1. The actual capacity estimation of WESS

Table 3 and Table 4 show the actual capacity measurements of WESS system. Ideally full constant discharges are needed but the operational window of the system is limited to 95% - 5% SOC. Therefore, the capacity tests were done between 95% and 5% SOC in May 2018 and between 90% and 10% SOC in Oct. 2019. The system was cycled with constant powers of 0.5MW($\sim$ 0.5C), 1MW($\sim$ 1C) and 2MW($\sim$ 2C), and tested 3 times for each power rate. Based on Eq. (6) the capacity was calculated by taking the integral of current and dividing this by 90% or 80%. These results show that the capacity degradation of the system over 16 months is negligible and the variations seen are likely measurement noise. Some results are higher than 1600 Ah and this is because actual cell capacities from manufacturer were higher than the nominal at installation.

Table 3: WESS Actual capacity calculations in May 2018

Results for 3 cycles	0.5C	1C	2C
Average Capacity (Ah)	1600.03	1614.43	1601.33
Standard deviation (Ah)	0.92	2.06	5.22
Relative standard deviation (%)	0.057	0.13	0.33

Table 4: WESS Actual capacity calculations in Oct.2019

Results for 3 cycles	0.5C	1C	2C
Average Capacity (Ah)	1601.03	1605.17	1598.9
Standard deviation (Ah)	1.17	5.2	1.7
Relative standard deviation(%)	0.073	0.324	0.106

4.2.2. Capacity estimation results with different datasets

Fig. 6 shows the capacity results of the same profiles in Fig. 5. Note that for these results the number of data points for each interval is tuned and fixed, and the noise on the current measurement is estimated based on empirical knowledge. The assumed capacity is 1600 Ah and the error borders refer to this value. In (a) and (b), the algorithm takes some time to converge because the estimated capacity was initialised as 0 (for worst-case demonstration), and the reason of no results at the beginning of (a) is that the SOC values were maintained at 50%, so there were no SOC variations, which is the denominator for calculating Q using Eq. (6). After that, these results are stable and mostly within the error borders. In comparison, the results in (c) show that the algorithm's performance is affected by the quality/type of data. There are several reasons for the worse results: a relatively flat SOC profile, this is because the error is relatively significant with small SOC variation, according to Eq. (6); sharp and short spikes in current data, which leads to inaccurate current integration due to sampling rate.

4.3. Data selection for online capacity estimation

The development of "online" capacity estimation is essential for the investigation of the degradation of large-scale battery systems during any different operational modes. For online estimation, the data is first obtained from time-

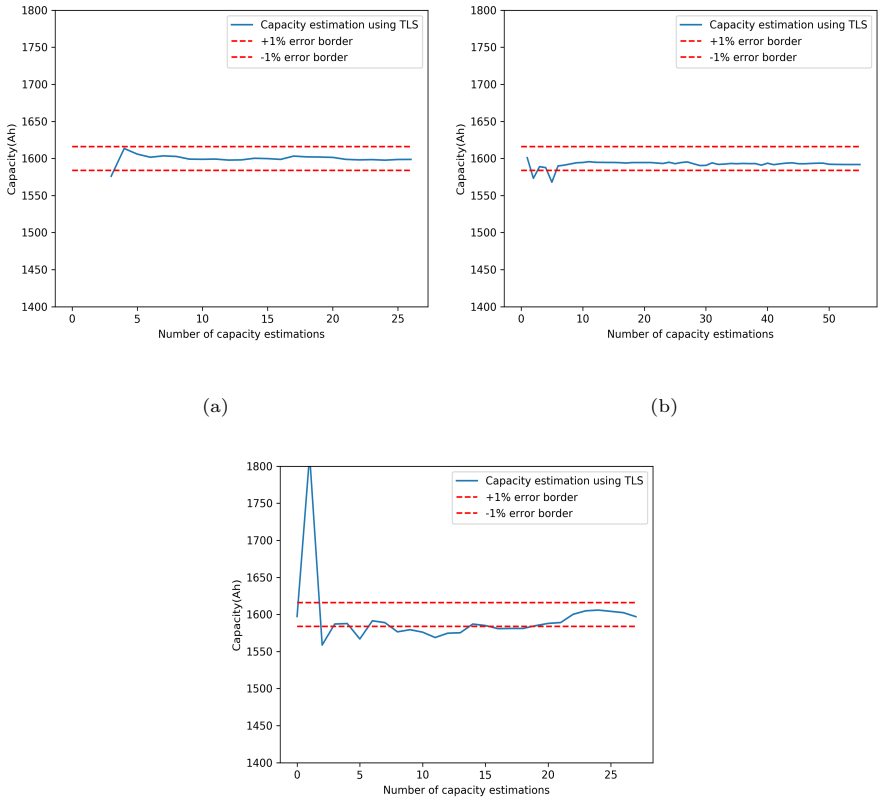


Fig. 6: Capacity estimation results of (a) constant cycling ($\sim$ 6.5 hours) and (b) mixed profile ($\sim$ 9 hours) and (c) dynamic frequency response ($\sim$ 7 hours). The red dotted lines show $\pm$1% error around the 1600 Ah assumed capacity.

series data, followed by the capacity estimation using the TLS algorithm. There are some restrictions for data to achieve online accurate capacity estimation:

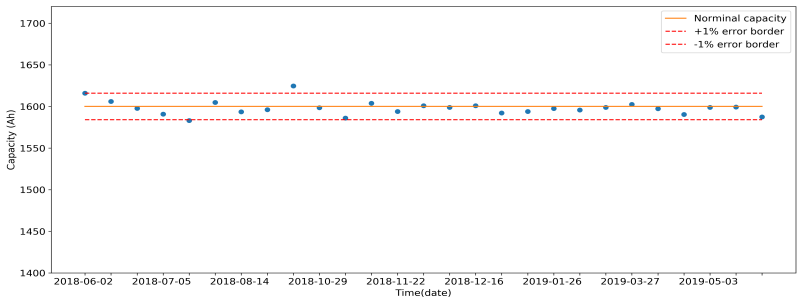
1. Continuous availability of significant variations in SOC data to maintain accuracy as discussed above.
2. A large interval size for calculation: The data is divided into a range of intervals with a size of m, this value can be different from the macro scale in the multi-scale DSPKF introduced above. This makes sure there are some SOC variations for every calculation. This value is set to 500 samples for the results shown in Fig. 6.
3. Sufficient data: the data length should not be less than a day so that the algorithm has enough time to converge.
4. No sharp, short spikes of current data: as discussed, such current data leads to inaccurate current integration.

A battery's capacity is not changing rapidly, so there is no need for estimating capacity frequently, which makes the selection of data that meets the above requirements possible.

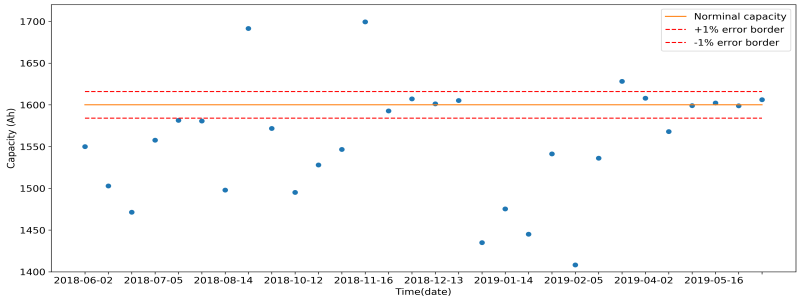
5. Comparison of using BMS SOC or DSPKF SOC for TLS capacity estimation

In this section a comparison for accuracy between using the BMS SOC and the DSPKF SOC as an input for the TLS capacity estimation algorithm. Fig. 7 (a) and (b) show the capacity estimation results using WESS operational data from 01/06/2018 to 01/06/2019, using the DSPKF SOC and the BMS SOC respectively. The dataset (current, BMS SOC, voltage and time) were divided into chunks (the time duration of every chunk is just over a week) to be processed. The DSPKF algorithm is ran on these sets of data to provide an estimate of SOC with the capacity parameter set to the nominal capacity (1600 Ah) of the system.

It can be seen that the errors of TLS capacity estimation results using the DSPKF SOC are mostly within the 1% error bounds of the measured capacity,



(a)



(b)

Fig. 7: Capacity results of one year using (a) DSPKF estimated SOC (b) BMS SOC

but the errors of using BMS SOC are much higher. Therefore, it could be concluded that replacing the BMS SOC with the DSPKF SOC can significantly improve the accuracy. However, it can be argued that this result is achieved when the system is relatively healthy using the nominal capacity as an initial capacity for SOC estimation, for a degraded system there would be significant error according to Eq. (3). When the battery system is degraded, the Q_{actual} in the equation will be larger than it actually is, this affects the accuracy of DSPKF SOC algorithm, thereafter decreasing the accuracy of TLS capacity estimation algorithm.

The practical and simple way to account for degradation is to use the Coulomb counting method. Despite stating that this is not good enough for accurate SOC estimation, accurate capacity information for DSPKF is not needed, since only a crude estimation is enough to ensure the SOC converges to a good enough value.

The multi-scale DSPKF algorithm introduced above makes the capacity correction based on SOC prediction possible. After every macro time duration, the SOC can be predicted using Coulomb counting with estimated capacity, and to be compared with the micro time scale estimated DSPKF SOC or the BMS SOC (reference SOC). Eq. (10) shows the calculation of predicted SOC [22].

$$SOC_{k,L} = SOC_{k,0} + \frac{T}{C_k^-} \sum_{j=0}^{L-1} -I_{k,j} \tag{10}$$

where $SOC_{k,0}$ is the SOC estimation at the beginning of a macro scale, k the number of micro SPKF estimations, T the sample rate in the micro SPKF, C_k^- the initialised or last estimated capacity, L the number of samples in every macro scale estimation and $SOC_{k,L}$ is the projected SOC. The correction of capacity is shown in the equation below:

$$C_k^+ = C_k^- + K(SOC_{k,L} - SOC_k) \tag{11}$$

where K is a gain, SOC_k the reference SOC, and C_k^+ is the corrected SOC. The purpose of using K is to accelerate the convergence since the SOC is a value between 0 and 1. K is tuned to control the correction speed, and is a

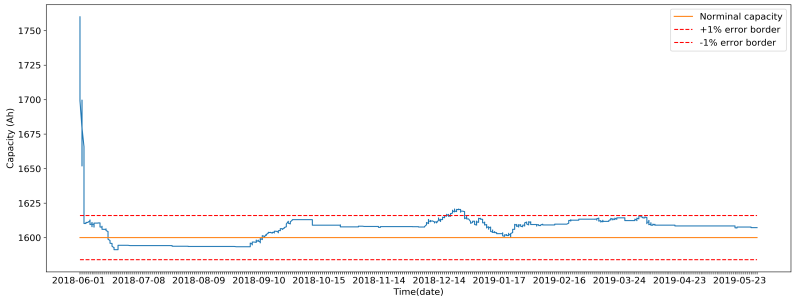


Fig. 8: Capacity tracking using capacity correction technique over one year of data

positive value when the system is charging and negative when the system is discharging. For example, assuming the last estimated capacity is larger than actual when the system is charging, according to Eq. (10), the projected SOC would be smaller than the micro SPKF estimated SOC or BMS SOC, so the SOC difference in the bracket in the equation above is a negative value, which multiplies a positive K. Thus, the last estimated capacity value will be decreased toward the actual capacity.

The results of capacity tracking over the year are illustrated in Fig. 8. The initial capacity was set as 1760 Ah, which is 10% higher than the actual capacity to simulate degradation. The errors are mostly within the 1% error bounds again for the year from 01/06/2018 to 01/06/2019. These capacity results can be provided regularly to update the DSPKF algorithm, creating a reliable input for the TLS algorithm; more accurate capacity estimation can then be obtained. Long-time results of this method are generally acceptable but the disadvantage is that the gain (K) is difficult to tune and it relies on the quality of the data. For example, if K is too large, the capacity could be over-corrected without good quality data, and if K is too small, this method requires a long time to converge to be within an acceptable range.

6. Conclusions

This work aims to provide methods for the state estimation of BESSs, which are essential for maximising their technical potential and return on investment. It has been proposed, and experimentally validated, that modelling a BESS as a single cell it is possible to achieve accurate estimation of SOC and capacity for the entire system. The advantage of these methods is that they respect the limits of data measurement accuracy and granularity in the real-world application as BESSs are scaled-up. However, it is clear that data selection is necessary, particularly for capacity estimation to achieve convergence and accurate results, and whilst selection criteria are presented in this paper, the techniques for optimal selection should be an area for future research. Another advantage of this methodology compared to alternative approaches where cell-level models are scaled-up, is that this approach is significantly less computationally demanding and requires minimal cell-level knowledge. For example, it is shown that either the OCV-SOC relationship measured at the system-level or at cell-level (and scaled appropriately) can be used, making this methodology suitable for situations where cell testing/data is not available. The overall methodology as presented in Fig. 1 has been implemented in real-time on a large-scale LTO BESS and demonstrated to provide reliable results. Future work will be required to assess the effectiveness of the capacity estimation with degradation on other types of BESSs and possible estimation of equivalent circuit parameters, especially the series resistor, as an output of the weight filter, for degradation modelling and end of life prediction.

Acknowledgements

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

References

[1] M. Kampa, E. Castanas, Human health effects of air pollution, Journal of The Electrochemical Society 151 (2008) 362– 367.

[2] O. Goksu, R. Teodorescu, C. L. Bak, F. Iov, P. C. Kjaer, Instability of wind turbine converters during current injection to low voltage grid faults and PLL frequency based stability solution, IEEE Transactions on Power Systems 29 (2014) 1683–1691.

[3] B. Xu, S. Member, Y. Shi, S. Member, Optimal battery participation in frequency regulation markets, IEEE Trans. Power Syst. 33 (6) (2018) 6715–6725.

[4] B. Gundogdu, S. Nejad, D. T. Gladwin, M. Foster, D. Stone, A battery energy management strategy for u.k. enhanced frequency response and triad avoidance, IEEE Trans. Ind. Electron. 65 (2018) 9509 – 9517.

[5] M. A. Hannan, M. S. H. Lipu, A. Hussain, 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. 78 (2017) 834–854.

[6] R. Xiong, L. Li, J. Badeda, J.Tian, Towards a smarter battery management system: A critical review on battery state of health monitoring methods, Journal of Power Sources 405 (2018).

[7] G. L. Plett, Efficient battery pack state estimation using Bar-Delta filtering, EVS24 International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium (2009).

[8] H. Dai, X. Wei, Z. Sun, J. Wang, W. Gu, Online cell soc estimation of li-ion battery packs using a dual time-scale kalman filtering for ev applications, Applied Energy 95 (2012) 227–237.

[9] R. Xiong, F. Sun, X. Gong, H. He, Adaptive state of charge estimator for lithium-ion cells series battery pack in electric vehicles, Journal of Power Sources 242 (2013) 699–713.

[10] W. Li, M. Rentemelster, J. Badeda, D. Jost, D. Schulte, Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation, Journal of Energy Storage 30 (2020).

[11] J. Hernández, F. Sanchez-Sutil, F. Muñoz-Rodríguez, Design criteria for the optimal sizing of a hybrid energy storage system in PV household-prosumers to maximize self-consumption and selfsufficiency, Energy 186 (2019).

[12] M. Gomez-Gonzalez, J. Hernández, D. Vera, F. Jurado, Optimal sizing and power schedule in pv household-prosumers for improving PV self-consumption and providing frequency containment reserve, Energy 191 (2020).

[13] J. Hernández, F. Sanchez-Sutila, F. Muñoz-Rodríguez, C. Baier, Optimal sizing and management strategy for PV household-prosumers with self-consumption/sufficiency enhancement and provision of frequency containment reserve, Applied Energy 277 (2020).

[14] K. S. Ng, C.-S. Moo, Y.-P. Chen, Y.-C. Hsieh, Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries, Applied Energy 86 (9) (2009) 1506–1511.

[15] W. Chang, The state of charge estimating methods for battery: a review, ISRN Appl Math (2013).
URL http://dx.doi.org/10.1155/2013/953792

[16] G. L. Plett, Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs - part 1. modeling and identification, J. Power Sources 134 (2) (2004) 262–276.

[17] 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 161 (2) (2006) 1356–1368.

[18] M. Nørgaard, N. K. Poulsen, O. Ravn, New developments in state estimation for nonlinear systems, Automatica 36 (2000) 1627– 1638.

[19] G. L. Plett, Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs part 2: Simultaneous state and parameter estimation, J. Power Sources 161 (2) (2006) 1369–1384.

[20] G. L. Plett, Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs, part 3: State and parameter estimation, J. Power Sources 134 (2) (2004) 277–292.

[21] Y. Zou, X. Hu, H. Ma, S. E. Li, Combined state of charge and state of health estimation over lithium-ion battery cell cycle lifespan for electric vehicles, J. Power Sources 273 (2015) 793–803.

[22] C. Hu, B. D. Youn, J. Chung, A multiscale framework with Extended Kalman Filter for lithium-ion battery SOC and capacity estimation, Applied Energy 92 (2012) 694–704.

[23] R. E. Kalman, A new approach to linear filtering and prediction problems, Journal of Basic Engineering 82 (1960) 35– 45.

[24] X. Hu, C. Zou, C. Zhang, Y. Li, Technological developments in batteries: A survey of principal roles, types, and management needs, IEEE Power Energy Mag 15 (5) (2017) 20–31.

[25] X. Hu, S. Li, Z. Jia, B. Egardt, Enhanced sample entropy-based health management of li-ion battery for electrified vehicles, Energy 64 (2014) 953–960.

[26] A. Farmann, W. Waag, A. Marongiu, D. U. Sauer, Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles, J. Power Sources 281 (2015) 114–130.

[27] G. L. Plett, Recursive approximate weighted total least squares estimation of battery cell total capacity, J. Power Sources 196 (4) (2011) 2319–2331.

[28] Toshiba, SCiB cells, https://www.scib.jp/en/product/cell.htm (2019).

[29] R. Xiong, F. Sun, Z. Chen, H. He, A data-driven multi-scale extended Kalman filtering based parameter and state estimation approach of lithium-ion polymer battery in electric vehicles, Applied Energy 113 (2014) 463–476.

[30] S. Nejad, D. T. Gladwin, D. A. Stone, A systematic review of lumped-parameter equivalent circuit models for real-time estimation of lithium-ion battery states, J. Power Sources 316 (2016) 183–196.

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

[32] M. Chen, G. A. Rincon-Mora, Accurate electrical battery model capable of predicting runtime and I–V performance, IEEE Transactions on Energy Conversion 21 (2006) 504– 511.

[33] S. Sivanandam, S. Deepa, Genetic Algorithms, Springer, Berlin Heidelberg, 2008.

[34] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, A fast and elitist multi-objective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.

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