Battery Identification Based on Real-World Data

Miao Zhang, *Student Member, IEEE,* Zhixin Miao, *Senior Member, IEEE,* Lingling Fan, *Senior Member, IEEE*

***Abstract*—In this paper, system identification is carried out for a 20 kWh battery using real-world measurements data. State-of-charge (SOC) and open-circuit voltage (V_{OC}) relationship will be obtained using least square estimation (LSE) non-linear regression. In addition, how to estimate SOC using current measurements and how to estimate the equivalent circuit's RC parameters are carried out using autoregressive exogenous (ARX) models. The respective ARX models are first derived. Estimation of the ARX coefficients is then carried out. Finally, parameter recovery is conducted to find out parameters with physical meanings, e.g., RC values. With the identified V_{OC} and SOC relationship and RC parameters, we built a simulation model in MATLAB/Simpowersystems. With the measured current data from the real-world as the input, the simulation model gives the terminal DC voltage as the output. This output is compared with the real-world DC voltage measurements data and the matching degree is satisfactory.**

***Index Terms*—state-of-charge (SOC), least-square-estimation (LSE), autoregressive exogenous (ARX) model, battery equivalent model, system identification, data analysis.**

I. Introduction

In recent years, batteries are used more and more. The target battery pack in this paper is located in St. Petersburg of Florida with the characteristics listed in TABLE I. It serves as an energy storage device. It is charged by Photovoltaic (PV) in the morning and discharged in the afternoon to mitigate electric power consuming. Fig. 1 shows the raw data obtained from meters, including measured terminal DC voltage (V_L), measured terminal DC current (I_L) and state-of-charge (SOC) in time-series. The time span is 22 days and sample time is 1 minute. The data extraction for analysis relies on data programming in Python.

TABLE I
Battery Main Characteristics

Rated Capacity	20 kWh
Rated Power	5 kW
Cell Rated Capacity	400 Ah

This paper has three objectives. The first one is to obtain open-circuit voltage (V_{OC}) to SOC relationship from voltage measurements and SOC data. The second one is to estimate SOC from current measurements. The third one is to estimate the equivalent circuit's RC parameters from the measured current and SOC.

There are at least two major systematic methods for battery system estimation: Kalman filter-based estimation and least-square estimation (LSE). For battery system identification, Kalman filter-based estimation, including Extended Kalman filter (EKF) [1], [2], Unscented Kalman filter (UKF) [3]–[5], have been widely used in state-of-charge (SOC) estimation and parameters identification. Kalman filter based method is a way to estimate the time-varying dynamic system with Gaussian noise. Kalman filter can be implemented online. On the other hand, LSE is chosen as a fast and efficient polynomial estimation method to identify battery systems in [6]–[8]. By approximating derivatives discrete data, discrete-time ARX models can be found [9], [10]. With an ARX model, a linear LSE problem can be formulated and the parameters of the ARX model can be estimated. In [9], [10], autoregressive exogenous (ARX) model are applied to generator system parameters identification. There are very few papers to estimate battery systems parameters using ARX model. Our paper fills the gap.

This paper is organized as follows. In Section II, V_{OC} and SOC relationship will be obtained using LSE non-linear regression. How to estimate SOC using current measurements and how to estimate the equivalent circuit's RC parameters are carried out using ARX model in Section III and IV, respectively. In Section V, with the identified V_{OC} to SOC relationship and RC values, we build a simulation model in MATLAB/Simpowersystems for validation. The conclusion is given in Section VI.

M. Zhang, Z. Miao and L. Fan are with Dept. of Electrical Engineering, University of South Florida, Tampa FL 33620. Email: zmiao@usf.edu

II. Estimation of Open Circuit Voltage and SOC Relationship

The relationship curve of open-circuit-voltage and SOC is usually used as a criterion to describe the battery health status. In [3], [8], [11]–[13], the experiment currents are kept constant so to get V_{OC} under different SOCs easily. However, the battery terminal current (I_L) and voltage (V_L) in our data vary with weather condition and power demand in real-time. The battery terminal voltage (V_L) equals the open-circuit-voltage (V_{OC}) when no current flows, as shown in Fig. 1 and pointed by the red arrows.

Normally, the open-circuit-voltage (V_{OC}) of a battery is influenced by SOC, working temperature T and the number of cycles C, as shown in (1).

$$V_{OC} = f(SOC, \quad T, \quad C) \tag{1}$$

In our case, the V_{OC} and SOC are acquired at each early morning when the battery environment temperature varies little according to the historic record. Also, considering 15 cycles in 22 days, the influence of cycle C can be neglected. Then (1) is simplified as (2). The triangle markers are used to represent the extracted 13 pairs of V_{OC} and SOC data in Fig. 2.

$$V_{OC} = f(SOC) \tag{2}$$

We can assume a target relationship function as a combination of exponential and polynomials. It is expressed in (3).

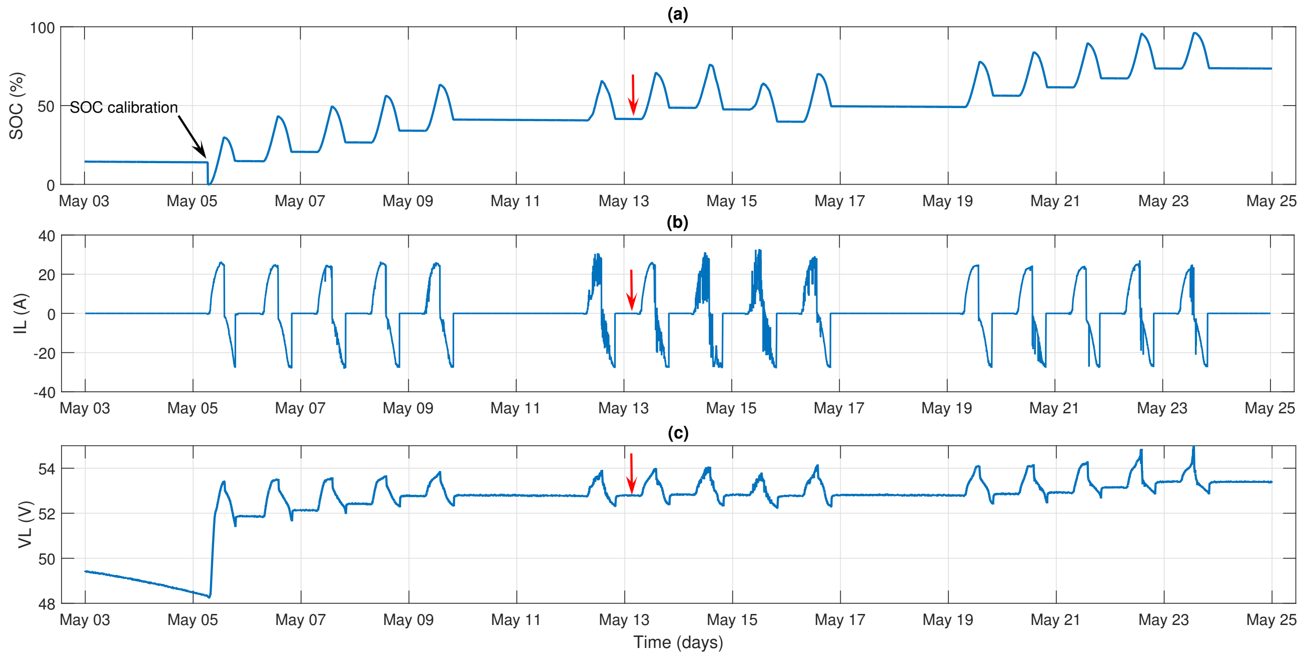

Fig. 1. (a) State-of-charge (SOC). (b) Battery terminal current measurements. (c) Battery terminal voltage measurements.

$$V_{OC} = a \cdot e^{b \cdot SOC} + p_1 \cdot SOC^7 + p_2 \cdot SOC^6 + p_3 \cdot SOC^5 + p_4 \cdot SOC^4 + p_5 \cdot SOC^3 + p_6 \cdot SOC^2 + p_7 \cdot SOC + p_8 \tag{3}$$

We aim to find the coefficients: $a, b, p_1, ...p_8$ that can best fit the curve to the given data points. The objective function is

$$\min_{a,b,p_1,\cdots,p_8} \sum_{i=1}^{n} (V_{OC}(a, b, p_1, \cdots, p_8) - V_{OCm})^2, \tag{4}$$

where V_{OC} is the estimation from SOC, V_{OCm} is the voltage measurements.

By applying curve fitting toolbox in MATLAB, we get the coefficients of (3) and a single-variable function is used to represent the curve, as shown in (5).

$$\begin{aligned} V_{OC} = & -4 \cdot e^{-0.3 \cdot SOC} + 9.431 \times 10^{-12} \cdot SOC^7 \\ & - 2.981 \times 10^{-9} \cdot SOC^6 + 3.541 \times 10^{-7} \cdot SOC^5 \\ & - 1.899 \times 10^{-5} \cdot SOC^4 + 3.965 \times 10^{-4} \cdot SOC^3 \\ & + 9.775 \times 10^{-4} \cdot SOC^2 - 0.08582 \cdot SOC + 52.32 \end{aligned} \tag{5}$$

This function is plotted in Fig. 2 and it is shown that the curve fits the data points very well.

III. SOC Estimation

In this Section, the general discrete-time ARX model structure will be first explained. SOC estimation and the current measurements (I_L) relationship will be converted to discrete time based on Coulomb counting method.

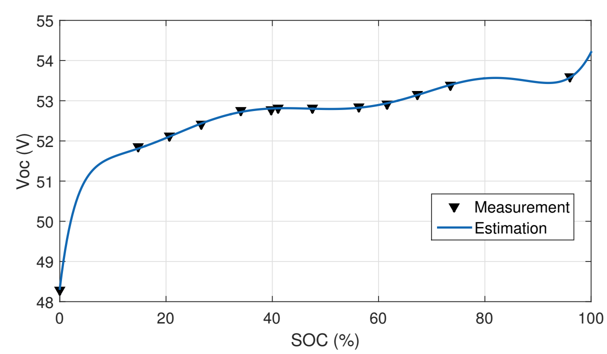

Fig. 2. SOC vs. V_{OC} measured points and estimated nonlinear relationship.

A. ARX Model Structure

A general polynomial ARX model structure can be expressed as equation (6):

$$A(z)y(k) = B(z)u(k - n_k) + e(k), \tag{6}$$

where

- $u(k)$ is the system inputs.
- $y(k)$ is the system outputs.
- n_k is the system delay.
- $e(k)$ is the White-noise system disturbance.
- $A(z)$ and $B(z)$ are polynomial with k_a and k_b orders respect to the backward shift operator z^{-1} and defined by the following equations:

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_{k_a} z^{-k_a} \tag{7}$$

$$B(z) = b_0 + b_1 z^{-1} + \cdots + b_{k_b-1} z^{-(k_b-1)} \tag{8}$$

Fig. 3 depicts the signal flow of an ARX model. Given time series no-delay ($n_k = 0$) measurements of the input and output, say from k step to N step, an overestimated problem can be formulated as (9) based on (6):

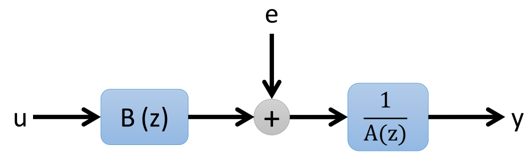

Fig. 3. ARX model signal flowchart.

$$\begin{bmatrix} y(k) \\ y(k+1) \\ \dots \\ y(N) \end{bmatrix} = \begin{bmatrix} y(k-1) & y(k-2) & \dots & y(k-k_a) & u(k) & u(k-1) & \dots & u(k-(k_b-1)) \\ y(k-2) & y(k-3) & \dots & y(k-k_a-1) & u(k-1) & u(k-2) & \dots & u(k-k_b) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y(N-1) & y(N-2) & \dots & y(N-k_a) & u(N) & u(N-1) & \dots & u(N-(k_b-1)) \end{bmatrix} \begin{bmatrix} -a_1 \\ \vdots \\ -a_{ka} \\ b_0 \\ b_1 \\ \vdots \\ b_{k_b-1} \end{bmatrix} + e(k) \tag{9}$$

B. ARX Model-based SOC Estimation

Theoretically, the state of charge is a relative quantity that describes the ratio of the remaining capacity to the nominal capacity of a battery. The Coulomb counting is a method developed from this concept, which estimates SOC by the measurements of total current accumulation. It can be given by

$$SOC = SOC_0 + \frac{\int \eta I_L dt}{C_N}, \tag{10}$$

where SOC_0 is the initial value of the SOC, η is the Coulombic efficiency, and C_N is the nominal capacity.

In our case, the battery had been fully discharged and experienced a long time of self-discharge. Shown in Fig. 1, I_L was kept 0 while V_L kept decreasing before May 5th. This is due to the effect of battery self-discharging. The SOC_0 was calibrated to zero on May 5th as shown in Fig. 1 (a) and pointed by the black arrow.

Convert (10) to discrete-time as:

$$SOC(k) = SOC(k-1) + \frac{\eta \Delta t}{C_N} I_L(k-1) \tag{11}$$

Assume $\frac{\eta \Delta t}{C_N}$ as a constant number b_1. From (11), the transfer function can be expressed by:

$$\frac{y}{u} = \frac{b_1 z^{-1}}{1 - z^{-1}} \tag{12}$$

Applying this to discrete-time ARX model, we can get the transfer function with order of [1 1 1]. This yields to

$$A(z)SOC(k) = B(z)I_L(k) + e(k), \tag{13}$$

where

$$A(z) = 1 - z^{-1}$$
$$B(z) = 0.004615 z^{-1}$$

With above discrete ARX transfer function, SOC estimation can be expressed as:

$$SOC(k) = SOC(k-1) + 7.69 \times 10^{-5} \Delta t \cdot I_L(k-1) + e(k) \tag{14}$$

where

$$SOC(0) = 0$$
$$\Delta t = 60\text{seconds}$$

Given I_L as input, we can compare ARX model based SOC simulation output with SOC raw data in Fig. 4. The Simulated ARX model output matches the raw data very well with a fitting degree of 98.59% and a mean-square error 0.06522.

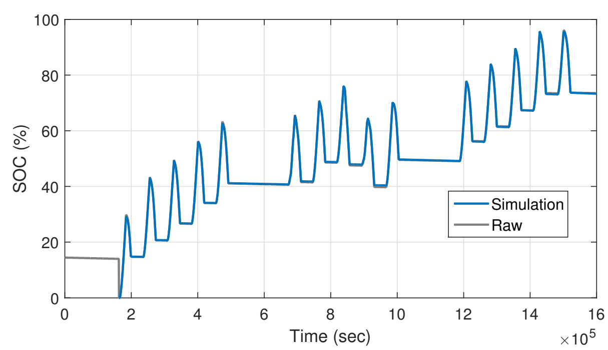

Fig. 4. Comparison of SOC ARX model simulation output and SOC raw data.

IV. Equivalent Circuit Parameters Identification

In this section, the battery equivalent circuit is first proposed. The dynamic equations of two RC branches are converted to discrete-time in z-domain. Finally, estimation of ARX Model coefficients are carried out and RC parameters recovery is conducted with physical meaning.

A. Equivalent Circuit Modeling

Among battery equivalent circuits, the Thevenin-based model is widely used in [3], [8], [11], [12], [14]–[17] since it can not only bridge SOC to open-circuit voltage, but also simulate the transient response of load changing. It consists of two parts. One part is open-circuit voltage(V_{OC}), which in function of state-of-charge (SOC) as shown in Eq. (5). V_{OC} is presented by a voltage-controlled voltage source in Fig. 5. Another part is the RC network, including one ohmic

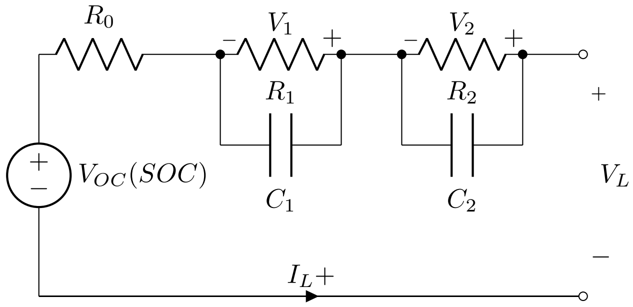

Fig. 5. Schematic of the battery equivalent circuit.

resistance R_0 and two paralleled RC branches (R_1, C_1 and R_2, C_2), are responsible for short-time and long-time constants of the step response. The proposed equivalent model is a trade off between accuracy and complexity.

B. Discretization of Dynamic Equations

As assumed in Fig. 5, I_L is the terminal current with a positive value in the charging process and negative value in the discharging process. Two RC parallel branches in proposed model can be expressed as following differential equations:

$$C_1 \frac{dV_1(t)}{dt} = \frac{-V_1(t)}{R_1} + I_L \tag{15a}$$

$$C_2 \frac{dV_2(t)}{dt} = \frac{-V_2(t)}{R_2} + I_L \tag{15b}$$

The state space model is:

$$\begin{bmatrix} \dot{V_1} \\ \dot{V_2} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{-1}{R_1C_1} & 0 \\ 0 & \frac{-1}{R_2C_2} \end{bmatrix}}_{\text{A}} \times \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \underbrace{\begin{bmatrix} \frac{1}{C_1} \\ \frac{1}{C_2} \end{bmatrix}}_{\text{B}} \times I_L \tag{16a}$$

$$V_L - V_{OC} = \underbrace{\begin{bmatrix} 1 & 1 \end{bmatrix}}_{\text{C}} \times \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \underbrace{R_0}_{\text{D}} \times I_L \tag{16b}$$

From (16a) and (16b) we have the expression as:

$$\dot{x} = Ax + Bu \tag{17a}$$

$$y = Cx + Du \tag{17b}$$

where $x = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$, $u = I_L$ is the input, and $y = V_L - V_{OC}$ is the output. V_{OC} is assumed as known and can be found using (5). A discrete-time form of (17a) is arranged as (18), where $k = 1, 2, 3 \ldots$

$$x(k+1) = x(k) + (Ax(k) + Bu(k)) \cdot h, \tag{18}$$

where h is time interval. Substitute $x(k+1)$ by $z \cdot x(k)$:

$$z \cdot x(k) = x(k) + (AH + I) \cdot x(k) + Bh \cdot u(k) \tag{19}$$

$$\longrightarrow x(k) = [zI - (Ah + I)]^{-1} Bh \cdot u(k) \tag{20}$$

For the output, substitute (20) into (17b):

$$\frac{y(k)}{u(k)} = C \cdot [zI - (Ah + I)]^{-1} Bh + D \tag{21}$$

The corresponding transfer function $G(z)$ of (21) is:

$$G(z^{-1}) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{22}$$

where a_1, a_2, b_0, b_1 and b_2 are the coefficients relate to RC parameters (23a)$\sim$(23e). In our measurements data, the time interval h is 60 seconds.

$$a_1 = \frac{h}{R_1C_1} + \frac{h}{R_2C_2} - 2 \tag{23a}$$

$$a_2 = -\frac{h}{R_1C_1} - \frac{h}{R_2C_2} + \frac{h^2}{R_1R_2C_1C_2} + 1 \tag{23b}$$

$$b_0 = R_0 \tag{23c}$$

$$b_1 = \frac{h}{C_1} + \frac{h}{C_2} + \frac{hR_0}{R_1C_1} + \frac{hR_0}{R_2C_2} - 2R_0 \tag{23d}$$

$$\begin{aligned} b_2 = R_0 &+ \frac{h^2}{R_1C_1C_2} + \frac{h^2}{R_2C_1C_2} + \frac{h^2R_0}{R_1R_2C_1C_2} \\ &- \frac{h}{C_1} - \frac{h}{C_2} - \frac{hR_0}{R_1C_1} - \frac{hR_0}{R_2C_2} \end{aligned} \tag{23e}$$

C. ARX Model-based RC Estimation

In this case, the setting of ARX model is assumed as follows:

$$A(z)y(k) = B(z)u(k) + e(k), \tag{24}$$

where

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2}$$
$$B(z) = b_0 + b_1 z^{-1} + b_2 z^{-2}$$

- Inputs: $u(k) = I_L$
- Outputs: $y(k) = V_L - V_{oc}$
- Sample time: $T_s = 60$ seconds
- Total time: $S = 19$ days
- Order of the polynomial $A(z)$: $k_a = 2$
- Order of the polynomial $B(z)$ + 1: $k_b = 3$
- Input-output delay: $n_k = 0$

Using Matlab identification toolbox we can solve ARX Model with 68.82% simulation focus and 0.0099 mean-square-error (MSE). The ARX model simulated output and $V_L - V_{OC}$ evaluated output are compared and shown in Fig. 6. And the transfer function we can get is:

$$G(z^{-1}) = \frac{0.009229 - 0.01419z^{-1}0.004967z^{-2}}{1 - 1.817z^{-1} + 0.8168z^{-2}} \tag{25}$$

Applying the 3rd, 10th and 16th days data into ($V_L - V_{oc}$) ARX model to validate the reliability of coefficients we got in (25). All these three groups coefficients are compared and listed in TABLE II, shown that the variation of coefficients is acceptable so (25) is reliable to identify equivalent circuit parameters.

Substitute the coefficients in (25) into the system of equations (23a)-(23e). We can get the solution of circuit RC parameters are $R_0 = 0.009188\Omega$, $R_1 = 0.0068\Omega$, $R_2 = 0.0140\Omega$, $C_1 = 7.926 \times 10^6 F$, $C_2 = 2.38 \times 10^4 F$. In TABLE III, the RC

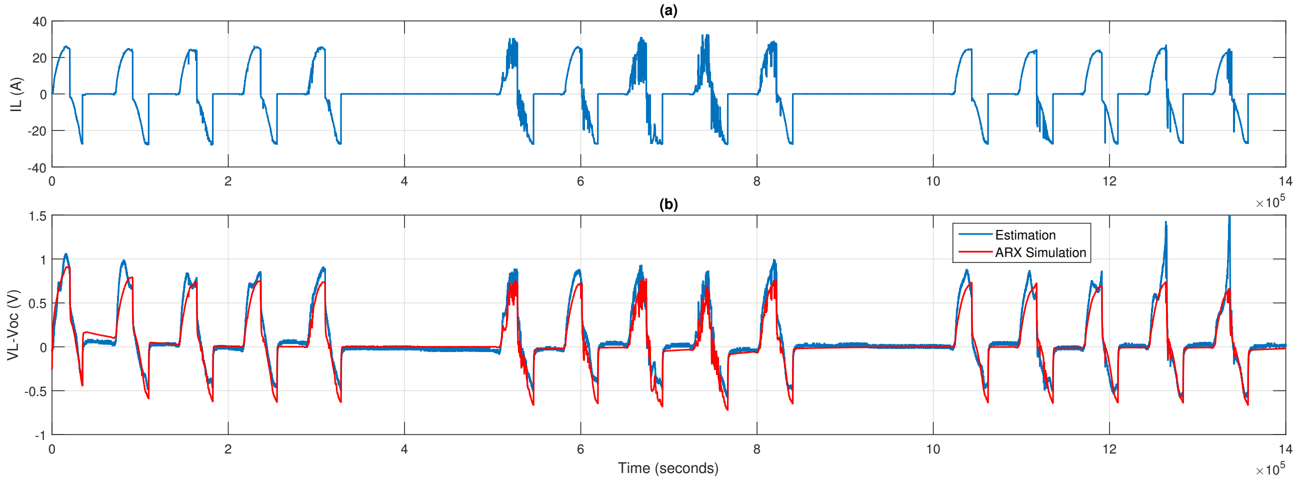

Fig. 6. (a) ARX model input I_L. (b) Comparison of $V_L - V_{OC}$ estimation and ARX model simulation

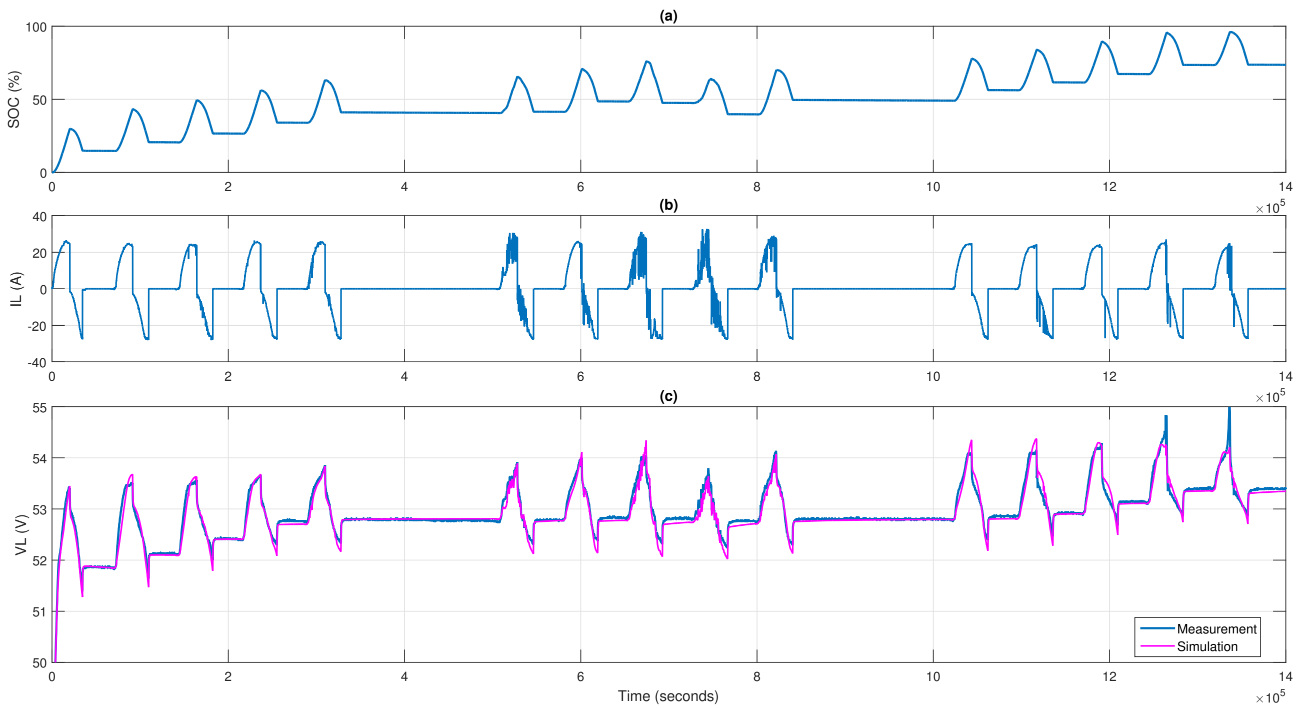

Fig. 7. (a) State-of-charge (SOC). (b) I_L input for current source. (c) Comparison of voltage measurements data and voltage simulation.

TABLE II
ARX MODEL COEFFICIENTS FROM DIFFERENT TIME PERIODS

a, b	19 days	3rd day	10th day	16th day	Max Deviation
a_1	-1.817	-1.879	-1.876	-1.887	3.8%
a_2	0.8186	0.8789	0.8759	0.8872	8.3%
b_0	0.009229	0.009308	0.008698	0.00833	9.7%
b_1	-0.01419	-0.01584	-0.01464	-0.01401	11.6%
b_2	0.0049676	0.006543	0.005945	0.005675	31.7%

TABLE III
RC PARAMETERS ESTIMATED FROM DIFFERENT TIME PERIODS

R, C	19 days	3rd day	10th day	16th day
$R_1(\Omega)$	0.0529	0.1320	0.0518	0.0495
$R_2(\Omega)$	0.0137	0.0127	0.0131	0.0162
$R_0(\Omega)$	0.009229	0.009308	0.008698	0.00833
$C_1(F)$	1.04×10^6	5.54×10^5	1.4454×10^6	6.74×10^5
$C_2(F)$	2.38×10^4	3.89×10^4	3.67×10^4	3.33×10^4

parameters which estimated from different periods are listed. Theoretically, the RC parameters are multi-variable functions of current, SOC, temperature and cycle number. It will lead the RC parameters variation without considering above all factors.

V. Validation Testbed

A simulation model, which shown in Fig. 8, was built to validate the proposed battery model performance. In this model, we set SOC and current (I_L) as two inputs. V_{OC} is in function of SOC in (5) as a voltage-controlled voltage source. I_L is the original measurements current acting as a current source on the terminal side. R,C values estimated from 19 days data is applied to R_0, R_1, R_2, C_1 and C_2. The terminal voltage in the Simulink model is obtained and compare it with the original battery measured voltage. All the inputs and outputs are presented in Fig. 7. By comparing the results, the mean squared error (MSE) is $0.01V$. The simulation measured DC voltage fit to raw measured DC voltage very well.

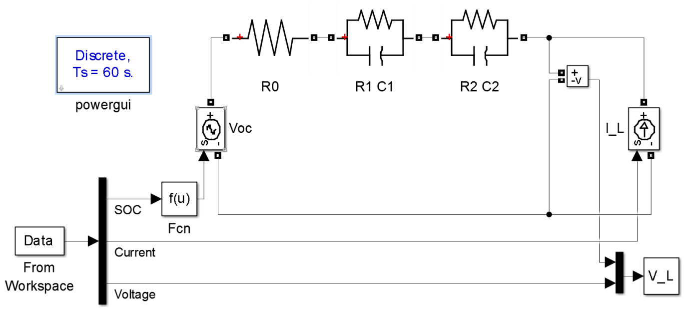

Fig. 8. Simulation testbed for validation.

VI. Conclusion

In this brief, system identification progress has been carried out for a 20 kW.h battery pack using real-world measurements data. V_{OC} to SOC relationship has been obtained by using least square estimation (LSE) non-linear regression. In addition, how to estimate SOC using current measurements and how to estimate the equivalent circuit's RC parameters were carried out using ARX model. Finally, with the identified V_{OC} to SOC relationship and RC parameters, we built a simulation model in MATLAB/Simpowersystems. With the measured current data from the real-world as the input, the simulation model gives the terminal DC voltage as the output. This output is compared with the real-world DC voltage measurements data and the matching degree is satisfactory.

References

[1] J. Lee, O. Nam, and B. Cho, "Li-ion battery soc estimation method based on the reduced order extended kalman filtering," *Journal of Power Sources*, vol. 174, no. 1, pp. 9–15, 2007.

[2] G. L. Plett, "Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 3. state and parameter estimation," *Journal of Power sources*, vol. 134, no. 2, pp. 277–292, 2004.

[3] H. He, R. Xiong, X. Zhang, F. Sun, and J. Fan, "State-of-charge estimation of the lithium-ion battery using an adaptive extended kalman filter based on an improved thevenin model," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1461–1469, 2011.

[4] J. Han, D. Kim, and M. Sunwoo, "State-of-charge estimation of lead-acid batteries using an adaptive extended kalman filter," *Journal of Power Sources*, vol. 188, no. 2, pp. 606–612, 2009.

[5] F. Sun, X. Hu, Y. Zou, and S. Li, "Adaptive unscented kalman filtering for state of charge estimation of a lithium-ion battery for electric vehicles," *Energy*, vol. 36, no. 5, pp. 3531–3540, 2011.

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

[7] X. Hu, F. Sun, Y. Zou, and H. Peng, "Online estimation of an electric vehicle lithium-ion battery using recursive least squares with forgetting," in *American Control Conference (ACC), 2011*. IEEE, 2011, pp. 935–940.

[8] H. He, X. Zhang, R. Xiong, Y. Xu, and H. Guo, "Online model-based estimation of state-of-charge and open-circuit voltage of lithium-ion batteries in electric vehicles," *Energy*, vol. 39, no. 1, pp. 310–318, 2012.

[9] B. Mogharbel, L. Fan, and Z. Miao, "Least squares estimation-based synchronous generator parameter estimation using pmu data," in *Power & Energy Society General Meeting, 2015 IEEE*. IEEE, 2015, pp. 1–5.

[10] Y. Xu, Z. Miao, and L. Fan, "Deriving arx models for synchronous generators," in *North American Power Symposium (NAPS), 2016*. IEEE, 2016, pp. 1–6.

[11] M. Chen and G. A. Rincon-Mora, "Accurate electrical battery model capable of predicting runtime and iv performance," *IEEE transactions on energy conversion*, vol. 21, no. 2, pp. 504–511, 2006.

[12] B. Schweighofer, K. M. Raab, and G. Brasseur, "Modeling of high power automotive batteries by the use of an automated test system," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 4, pp. 1087–1091, 2003.

[13] A. Szumanowski and Y. Chang, "Battery management system based on battery nonlinear dynamics modeling," *IEEE transactions on vehicular technology*, vol. 57, no. 3, pp. 1425–1432, 2008.

[14] L. Xu, Z. Miao, and L. Fan, "Control of a battery system to improve operation of a microgrid," in *Power and Energy Society General Meeting, 2012 IEEE*. IEEE, 2012, pp. 1–8.

[15] V. Johnson, "Battery performance models in advisor," *Journal of power sources*, vol. 110, no. 2, pp. 321–329, 2002.

[16] C. Zhang, J. Liu, S. Sharkh, and C. Zhang, "Identification of dynamic model parameters for lithium-ion batteries used in hybrid electric vehicles," 2010.

[17] L. Gao, S. Liu, and R. A. Dougal, "Dynamic lithium-ion battery model for system simulation," *IEEE transactions on components and packaging technologies*, vol. 25, no. 3, pp. 495–505, 2002.