Gaussian process regression-based modelling of lithium-ion battery temperature-dependent open-circuit-voltage

C. Huang and L. Wang✉

Open-circuit-voltage (OCV) plays a significant role in state-of-charge (SOC) estimation for lithium-ion batteries. The slight difference in OCV at various temperatures can result in a large deviation of SOC estimation. In this Letter, a novel model based on Gaussian process regression is proposed to describe the sophisticated relationship among the OCV, SOC, and temperature. To validate the effectiveness of the proposed model, a comprehensive comparison with widely considered benchmarking OCV models is conducted. Experiment results demonstrate the proposed model can provide the most accurate prediction of OCV compared with benchmarking models.

Introduction: In the automotive industry, electric vehicles (EVs) have attracted increasing attention and achieved ever better development as an important technology to reduce the greenhouse gas emission and the consumption of natural resources [1]. The battery system, as one of the key parts in EVs, plays a crucial role in determining the efficiency, reliability, and safety of the EVs systems. Due to the demanding driving operations, a battery management system (BMS) is required to guarantee the good performance of the battery. An accurate estimation of the state of charge (SOC) is one of the main functions for a BMS [2]. The SOC quantifies the remaining charge of the battery at the current cycle, which indicates the remaining available range for EVs.

The model-based approaches have been widely applied to SOC estimation [3–5]. In these approaches, the accuracy of SOC estimation highly depends on the open-circuit-voltage (OCV) model which relates SOC to OCV. Due to the relative flat OCV curve over a wide SOC especially for lithium iron phosphate ($LiFePO_4$) batteries, a small error on the inferred OCV will lead to a large deviation in SOC estimation. Several models including the look-up table and analytical models have been developed to implement the OCV–SOC function in a BMS. Analytical models demonstrate advantages in computational efficiency and convenience of analysis over the method of look-up table [6]. However, these models are often constructed at a certain temperature. Studies in [7] illustrate that the slight difference among OCV curves at various temperatures can cause significant errors in SOC estimation. EVs always operate under dynamic environmental conditions including varying temperatures. Hence, an accurate OCV model considering dependence on temperature is crucial to SOC estimation for EVs.

In this study, a temperature-dependent OCV model is developed relying on the Gaussian process (GP) regression (GPR). With this model, it is easy to make inferences of OCV at other temperatures. To validate the effectiveness of the proposed model, a comprehensive comparison with widely considered benchmarking models will be conducted.

Data description: The OCV–SOC curves at various temperatures over the whole SOC range from 0 to 100% are depicted in Fig. 1*a*, while the Fig. 1*b* clearly illustrates the significance of the accuracy of inferred OCV to SOC estimation in the flat region. The utilised data in Fig. 1 is collected from the Center for Advanced Life Cycle Engineering (http://www.calce.umd.edu/batteries/data.htm), University of Maryland, USA. The OCV–SOC experiments with A123 cell at various temperatures are described in [7]. In fact, the OCV in the charging process is higher than in the discharging process due to a hysteresis phenomenon. In this Letter, the OCV shown in Fig. 1 is defined as the average value of charging and discharging data, and we concentrate on the data in the middle SOC range from 10 to 90% where the OCV model is the most useful as discussed in [6]. The data points are divided into two groups. The first group includes data points at 0, 10, 30, and 40°C. In this group, 75% data points at each temperature are utilised for model development, while the remaining 25% data points are utilised for model validation. The second group includes data points at 20°C, which is considered as the test dataset to verify model accuracy in predicting OCV at a blind temperature.

GPR-based modelling of temperature-dependent OCV: To accurately model the sophisticated relationship among the OCV, SOC, and temperature (T), a GPR [8] model is applied. A GPR model relates the response to inputs through (1):

$$y = f(\boldsymbol{x}) + \xi \tag{1}$$

where y refers to OCV and $\boldsymbol{x} = [\text{SOC}, T]$. The ξ is an additive Gaussian noise with mean zero and variance σ^2. The $f(\boldsymbol{x})$ is latent variables from a GP which is a collection of random variables, and any finite number of such variables has a joint Gaussian distribution. A GP is specified by its mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ as $f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$. To simplify the notation, the mean function is usually assumed to be zero and the response variable is normalised to have zero mean.

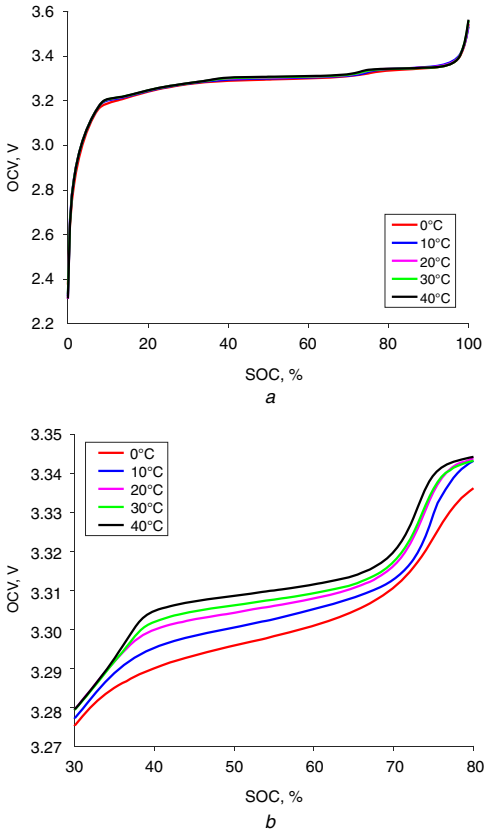

Fig. 1 *OCV–SOC curves at various temperatures*
a SOC from 0 to 100%
b SOC from 30 to 80%

In (1), the model response is a linear combination of Gaussian variables. Consider the training dataset $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{y})$, where $\boldsymbol{X} = \{\boldsymbol{x}_i|_{i=1}^N\}$ and $\boldsymbol{y} = \{y_i|_{i=1}^N\}$, then $\boldsymbol{y}$ also follows Gaussian distribution:

$$\boldsymbol{y} \sim N\big(0, K(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}\big) \tag{2}$$

where $K(\boldsymbol{X}, \boldsymbol{X})$ is the covariance matrix with $K[i, j] = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{I}$ is an identity matrix. When a new input, $\boldsymbol{x}^*$, is given, the joint distribution of the response, y^*, can be written as

$$\begin{bmatrix} \boldsymbol{y} \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I} & K(\boldsymbol{X}, \boldsymbol{x}^*) \\ K(\boldsymbol{x}^*, \boldsymbol{X}) & k(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{bmatrix}\right) \tag{3}$$

According to the theory of the joint Gaussian distribution, the prediction of y^* can be made by

$$y^* | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}^* \sim N\Big(\mu, \sum\Big) \tag{4}$$

where $\mu = K(\boldsymbol{x}^*, \boldsymbol{X})\big[K(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}\big]^{-1} \boldsymbol{y}$ and $\sum = k(\boldsymbol{x}^*, \boldsymbol{x}^*) - K(\boldsymbol{x}^*, \boldsymbol{X})\big[K(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}\big]^{-1} K(\boldsymbol{X}, \boldsymbol{x}^*)$.

In this study, the squared exponential covariance function (5) is considered. Model parameters including σ, σ_f, and l are estimated by maximising the logarithm marginal likelihood function on the

training dataset

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma_f^2 \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2l^2}\right) \quad (5)$$

Benchmarking models: To validate the effectiveness of the proposed model in OCV prediction considering the dependence on temperature, the proposed model is compared with following widely considered benchmarking models.

Model 1 [6]:

$$y = k_0 + k_1 \frac{1}{1 + e^{\alpha_1(x-\beta_1)}} + k_2 \frac{1}{1 + e^{\alpha_2(x-\beta_2)}} + k_3 \frac{1}{1 + e^{\alpha_3(x-1)}} + k_4 \frac{1}{1 + e^{\alpha_1 x}} + k_5 x \quad (6)$$

Model 2 [3]:

$$y = k_0 - k_1/x - k_2 x + k_3 \ln x + k_4 \ln(1 - x) \quad (7)$$

Model 3 [9]:

$$y = k_0 + k_1(1 - e^{-\alpha_1 x}) + k_2\left(1 - e^{-\alpha_2/(1-x)}\right) + k_3 x \quad (8)$$

Model 4 [10]:

$$y = k_0 + k_1 e^{-\alpha_1 x} + k_2 x + k_3 x^2 + k_4 x^3 \quad (9)$$

Model 5 [11]:

$$y = k_0 + k_1 x + k_2 x^2 + k_3 x^3 + k_4 x^4 + k_5 x^5 + k_6 x^6 \quad (10)$$

where x and y refer to SOC and OCV, respectively. In these models, k_0 to k_6, α_1 to α_3, and β_1 and β_2 are model parameters obtained by minimising the mean-squared error on the training dataset.

Experiment results: Computational results including the mean absolute error (MAE), root-mean-squared error (RMSE), and maximum absolute error (Max.AE) in mV are provided in Table 1. The results of benchmarking models (M1–M5) on the test dataset are obtained by linear interpolation with respect to temperature as the benchmarking models are constructed at each temperature separately.

Table 1: Results of different models

Errors (mV)		M1	M2	M3	M4	M5	Proposed
Train	MAE	1.86	3.61	21.65	15.21	1.26	**0.28**
	RMSE	2.65	4.57	25.21	21.18	1.72	**0.42**
	Max.AE	9.28	19.17	62.38	70.66	6.50	**1.88**
Val	MAE	2.22	4.07	21.61	15.14	1.54	**0.33**
	RMSE	3.21	5.43	25.55	20.34	2.08	**0.55**
	Max.AE	9.14	16.86	63.46	61.63	6.66	**3.42**
Test	MAE	2.52	4.01	20.33	20.73	1.77	**0.83**
	RMSE	3.40	5.29	24.39	25.19	2.29	**1.00**
	Max.AE	9.82	16.5	62.06	60.44	7.31	**3.20**

The lowest values are in bold.

It is observable in Table 1 that the proposed GPR-based OCV model performs much better than the widely considered benchmarking models in terms of MAE, RMSE, and maximum absolute error on the training, validation, and test datasets. The modelling OCV–SOC curve on the test dataset by the proposed model is depicted in Fig. 2. It is observable that the predicted curve matches well with the measured one.

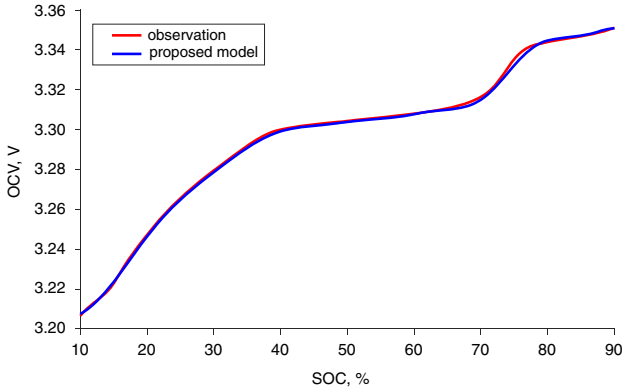

Fig. 2 *OCV–SOC curve by the proposed model on the test dataset*

Conclusions: A novel approach based on GPR was proposed to model the OCV–SOC function considering the impacts of temperature. Compared to the widely considered benchmarking models, the proposed model performed best with lowest errors. This Letter supports that the GPR-based OCV model is applicable for the inference of OCV at high accuracy.

Submitted: *2 June 2017* E-first: *24 July 2017*
doi: 10.1049/el.2017.2136
One or more of the Figures in this Letter are available in colour online.

C. Huang and L. Wang (*Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong*)

✉ E-mail: long.wang@my.cityu.edu.hk

References

1 Wesseling, J.H., Faber, J., and Hekkert, M.P.: 'How competitive forces sustain electric vehicle development', *Technol. Forecast. Soc. Change*, 2014, **81**, pp. 154–164
2 Elsayed, A.T., Lashway, C.R., and Mohammed, O.A.: 'Advanced battery management and diagnostic system for smart grid infrastructure', *Trans. Smart Grid*, 2016, **7**, (2), pp. 897–905
3 Plett, G.L.: 'Extended Kalman filtering for battery management systems of Lipb-based Hev battery packs – Part 2. Modeling and identification', *J. Power Sources*, 2004, **134**, (2), pp. 262–276
4 Aung, H., Low, K.S., and Goh, S.T.: 'State-of-charge estimation of lithium-ion battery using square root spherical unscented Kalman filter (Sqrt-UKFST) in nanosatellite', *Trans. Power Electron.*, 2015, **30**, (9), pp. 4774–4783
5 Wang, Y.J., Zhang, C.B., and Chen, Z.H.: 'On-line battery state-of-charge estimation based on an integrated estimator', *Appl. Energy*, 2017, **185**, pp. 2026–2032
6 Weng, C.H., Sun, J., and Peng, H.: 'A unified open-circuit-voltage model of lithium-ion batteries for state-of-charge estimation and state-of-health monitoring', *J. Power Sources*, 2014, **258**, pp. 228–237
7 Xing, Y., He, W., Pecht, M., *et al.*: 'State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures', *Appl. Energy*, 2014, **113**, pp. 106–115
8 Rasmussen, C.E., and Williams, C.K.I.: 'Gaussian processes for machine learning' (MIT Press, Cambridge, MA, USA, 2006)
9 Hu, Y., Yurkovich, S., Guezennec, Y., *et al.*: 'Electro-thermal battery model identification for automotive applications', *J. Power Sources*, 2011, **196**, (1), pp. 449–457
10 Chen, M., and Rincon-Mora, G.A.: 'Accurate electrical battery model capable of predicting runtime and IV performance', *Trans. Energy Convers.*, 2006, **21**, (2), pp. 504–511
11 Szumanowski, A., and Chang, Y.H.: 'Battery management system based on battery nonlinear dynamics modeling', *Trans. Veh. Technol.*, 2008, **57**, (3), pp. 1425–1432