Letter

Knowledge-Guided Data-Driven Model With Transfer Concept for Battery Calendar Ageing Trajectory Prediction

Kailong Liu, *Member, IEEE*, Qiao Peng, Remus Teodorescu, *Fellow, IEEE*, and Aoife M. Foley, *Senior Member, IEEE*

Dear Editor,

Lithium-ion (Li-ion) battery has become a promising source to supply and absorb energy/power for many energy-transportation applications. However, Li-ion battery capacity would inevitably degrade over time, making its related ageing prediction necessary. This letter presents effective battery calendar ageing trajectory prediction by deriving a knowledge-guided data-driven model with transfer concept. More specifically, this data-driven model is based on the support vector regression (SVR) technology. To ensure highly-accurate prognostics of battery calendar ageing trajectory under witnessed conditions, a knowledge-guided kernel is first developed by coupling the mechanism and empirical knowledge elements of battery storage temperature, state-of-charge (SoC), and time. To improve model's generalization ability under unwitnessed conditions, the knowledge-guided data-driven model is then equipped with transfer concept by adding a classical Gaussian kernel for all inputs. A well-rounded real battery ageing dataset under eight different storage conditions is collected to evaluate the performance of developed model. Results illustrate that this knowledge-guided battery ageing trajectory prediction model presents satisfactory accuracy for witnessed conditions with R^2 over 0.98. After using only 20% starting capacity point to tune its transfer part, it can also generalize well for unwitnessed conditions with R^2 over 0.97, further heavily reducing the required ageing experimental time and cost.

Related work: Significant studies have been devoted to achieving reasonable battery calendar ageing trajectory prognostics recently [1]. They result in the three main types of models: physics-based, semi empirical-based, and data driven-based ones.

Physics-based models use the complicated partial differential equations (PDEs) to consider battery electrochemical mechanisms, which are able to accurately describe battery ageing dynamics [2]. However, due to the involved complex PDEs, they are generally time-consuming and difficult to be parametrized, making this type of model become difficult to be adopted in real-time applications.

Semi empirical-based models present the advantages of simple structure and being compact to be parametrized, which have been widely utilized in real battery calendar ageing trajectory prediction applications. These models generally belong to the open-loop type, further presenting poor generalization for unwitnessed storage cases.

With the rapid development of artificial intelligence and data science, data-driven models have been also widely utilized to the field of battery ageing trajectory predictions. As summarized in the related reviews [3], [4], on the one hand, after developing proper data-driven models based on various machine learning technologies, satisfactory battery ageing trajectory prediction results can be achieved for the cases that are similar to the training cases (witnessed conditions). On the other hand, their generalization ability and the costs of related experimental ageing tests to generate necessary training data have much room for improvement. This study focuses on making such important improvements.

Problem statement: Data-driven method has become a promising tool in the battery ageing prediction area owing to its merits of being flexible and strong nonlinear fitting capability. General data-driven models require test data present a similar nature as the training data. However, the nature of real battery ageing data can be different from that of training cases, leading to that the data-driven models owning accurate results in training conditions are also difficult in presenting high performance under unwitnessed conditions. This phenomenon is named as the conceptual drift issue in data-science, which is a challenge and limited attention has been paid to battery ageing prognostics [3]. In addition, to generate suitable data for model training, battery ageing experiment is required but usually laborious and time-consuming (several months to years). In this context, it is challenging but still vital to develop an efficient data-driven model which is trained by the easily-collected data and is able to perform good generalization ability under different storage conditions [4].

Calendar ageing experiments and dataset: To generate proper battery ageing data for model development and performance exploration, battery calendar ageing experiments under various storage cases are carried out in a real battery ageing platform, as shown in Fig. 1. To be specific, the platform contains three main parts: a Neware BTS-8000 battery charger to charge or discharge batteries for maintaining their storage SoCs, a thermal chamber to maintain battery storage temperatures, and a PC to collect and process ageing data. Here, the tested NCM622 battery presents a nominal capacity of 50 Ah.

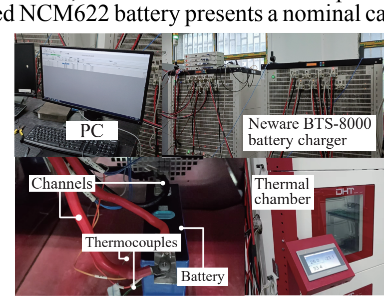

Fig. 1. Battery calendar ageing experiment platform.

In this study, eight different cases (two storage temperature and four storage SoC levels) covering the most storage conditions of this type of battery are explored with the detailed information illustrated in Table 1. Here, the SoC levels include 20%, 50%, 70% and 95%, while the temperature levels include 25℃ and 45℃. The corresponding battery calendar ageing data for each storage case are all collected by using thirteen capacity check-ups that occur once every 720 hours [h]. In the check-up process of each battery cell, the temperature chamber would be first set to 2℃. A constant-current-constant-voltage (CCCV) pattern is then performed to fully charge cell until the charging current in the constant-voltage (CV) stage decreases to C/20. After 2 hours resting period, a constant-current (CC) pattern with 1/3C is then utilized to fully discharge cell to the cut-off voltage of 2.75 V. Afterwards, the reference capacity will be calculated by using classical Coulomb counting approach.

Following the designed ageing experiments, eight battery capacity ageing series covering 8640 storage hours [h] are collected. Without the loss of generality, battery state-of-health (SoH) is utilized in all cases of this study to reflect battery ageing status through dividing battery actual capacity value by its rated capacity value, as recommended by [1]. In real applications, higher storage SoCs and temperatures are easy to increase the growth of solid electrolyte interface (SEI) film on the anode of Li-ion batteries, further causing battery capacity to degrade faster [5]. It thus becomes meaningful to develop a suitable model that can be just trained by using partial accelerated ageing data under high SoCs and temperatures (i.e., Cases 5−8 in Table 1), while the developed model is capable of generalizing well and conveniently to other unwitnessed storage conditions (i.e., Cases 1−4 in Table 1). One significant benefit of doing this is that the corresponding experimental time and costs of generating data for model

Corresponding author: Qiao Peng.

Citation: K. L. Liu, Q. Peng, R. Teodorescu, and A. M. Foley, "Knowledge-guided data-driven model with transfer concept for battery calendar ageing trajectory prediction," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 1, pp. 272–274, Jan. 2023.

K. L. Liu was with the Warwick Manufacturing Group, University of Warwick, Coventry CV4 7AL, UK. He is now with the School of Control Science and Engineering, Shandong University, China (e-mail: kliu02@qub.ac.uk).

Q. Peng and A. M. Foley are with the Queen's University Belfast, Belfast BT7 1NN, UK (e-mail: qpeng01@qub.ac.uk; a.foley@qub.ac.uk).

R. Teodorescu is with the Department of Energy Technology, Aalborg University, Aalborg 9220, Denmark (e-mail: ret@et.aau.dk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2023.123036

Table 1. Battery Ageing Matrix for Different Storage SoCs and Temperatures

SoC\Temperature	25℃	45℃
20%	Case 1	Case 2
50%	Case 3	Case 4
70%	Case 5	Case 6
95%	Case 7	Case 8

training will be heavily reduced.

Proposed knowledge-guided data-driven model: To achieve the abovementioned tasks, the proposed model is an improved SVR equipping battery ageing knowledge information and transfer concept. In theory, SVR is a kernel-based machine learning tool presenting good capability to handle the local minimum and overfitting issues. Supposing a dataset $\mathrm{D} = \{x_d, y_d\}_{d=1}^N$, N stands for the sample number, while x_d and y_d are the dth input vector and corresponding output, respectively. SVR can capture the relations between input and output by a nonlinear mapping Φ as

$$SVR(x) = \omega \cdot \Phi(x) + c \tag{1}$$

where x is the input, ω and c are its weight and bias. Then, the slack variables $\{\xi_d\}_{d=1}^N$ and $\{\xi_d^*\}_{d=1}^N$ can be formulated by

$$\min R(\omega, c, \xi) = \frac{1}{2}\|\omega\|^2 + P\sum_d^N \xi_d + \xi_d^* \tag{2}$$

$$\text{s.t.} \quad \begin{cases} y_d - \omega \cdot \Phi(x) - c \le \varepsilon + \xi_d \\ \omega \cdot \Phi(x) + c - y_d \le \varepsilon + \xi_d^* \\ \xi_d, \quad \xi_d^* \ge 0 \end{cases} \tag{3}$$

where P denotes a penalty factor, $\varepsilon (\varepsilon > 0)$ represents a maximum error allowed by prediction. After introducing the Lagrangian multipliers and kernel functions, it can be further transformed as

$$\begin{aligned} \max R\left(\alpha_d^*, \alpha_d\right) = & -\frac{1}{2}\sum_{d,e}^{N}\left(\alpha_d^* - \alpha_d\right)\left(\alpha_e^* - \alpha_e\right)\Phi(x_d)\Phi(x_e) \\ & -\sum_d^N \alpha_d (y_d + \varepsilon) + \sum_d^N \alpha_d^* (y_d - \varepsilon) \end{aligned} \tag{4}$$

with

$$\text{s.t.} \quad \begin{cases} \sum_d^N \left(\alpha_d^* - \alpha_d\right) = 0 \\ 0 \le \alpha_d, \quad \alpha_d^* \le P, \quad d = 1, \ldots, N \end{cases} \tag{5}$$

where α_d and α_d^* are Lagrangian multipliers. Then, the SVR-based nonlinear mapping could be derived by the Mercer's theorem as

$$SVR(x) = \sum_d^N (\alpha_d - \alpha_d^*) K(x_d, x) + c \tag{6}$$

where $K(x_d, x) = \Phi(x_d)\Phi(x)$ is the kernel function, which must be carefully designed to meet requirements of different applications.

Data-driven model structure: Based upon SVR, a data-driven model is designed to predict future battery calendar ageing trajectory. Fig. 2 details the structure of this data-driven model, which contains a knowledge-guided kernel part and a transfer part. Four items including battery capacity at the current time ($Cap(t)$), future storage time period (Δt_{fut}), storage temperature (T_{sto}) and storage SoC (SoC_{sto}) are inputs of this model, while battery capacity at the future time point ($Cap(t + \Delta t_{fut})$) is the output.

Temperature effect knowledge: In real battery applications, Arrhenius law [6] is usually adopted to describe the effects of temperature on the side reaction of battery in an exponential form, as illustrated by

$$f_B(T) = a_B \cdot \exp(-E_A/RT) \tag{7}$$

where a_B is a weight parameter. R is the ideal gas constant. E_A is the activation energy of side reactions. T is the ambient temperature.

To make use of the mechanism knowledge within the Arrhenius equation, the component $k_{Tst}(x_T, x_T')$ related to battery storage temperature T_{sto} is created by using the similar exponential form as

$$k_{Tst}\left(x_T, x_T'\right) = \sigma_{Tst} \cdot \exp\left(-\frac{1}{l_{Tst}}\left\|\frac{1}{x_T} - \frac{1}{x_T'}\right\|\right) \tag{8}$$

where σ_{Tst} and l_{Tst} are related hyper-parameters. According to this component, the effects related to the difference between various storage temperature cases x_T and x_T' are explained in an isotropic form. In this context, temperature impact knowledge is successfully coupled into the knowledge-guided kernel function.

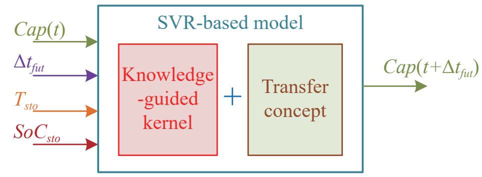

Fig. 2. Derived data-driven model structure.

SoC effect knowledge: Next, a proper component is created to take the knowledge of the storage SoC effect into account. In real battery storage applications, a higher storage SoC will generally lead to a larger degradation rate of calendar ageing. According to the related research [4], the effects of storage SoC on battery ageing present a linear relation. That is, the knowledge of the storage SoC effect could be described through a linear equation. In this context, to effectively involve this empirical knowledge of SoC effect, the component $k_{SoC}(x_{SoC}, x_{SoC}')$ related to battery storage SoC within the knowledge-guided kernel is created by using a linear form as

$$k_{SoC}\left(x_{SoC}, x_{SoC}'\right) = \sigma_{SoC} \cdot x_{SoC}^T x_{SoC}' + c_{SoC} \tag{9}$$

where σ_{SoC} and c_{SoC} are related hyper-parameters. In theory, a kernel with a linear form belongs to none stationary kernel, which can reduce the corresponding computational effort [7].

Time effect knowledge: On the basis of battery ageing knowledge, the effect of time on battery degradation is described by a power function as t^z (here t is storage time, z is power factor). In this context, to consider time effect knowledge into the kernel, a related component $k_{\Delta t}(x_{\Delta t}, x_{\Delta t}')$ related to Δt_{fut} is derived as

$$k_{\Delta t}(x_{\Delta t}, x_{\Delta t}') = \left(x_{\Delta t}^T x_{\Delta t}'\right)^z. \tag{10}$$

Coupled knowledge-guided kernel: After formulating the components for temperature, SoC, and time based on the mechanism and empirical knowledge of battery calendar ageing, a knowledge-guided kernel can then be designed. To also capture the battery capacity term $Cap(t)$ in Fig. 2, a classical Gaussian kernel is adopted. In theory, the matrix of kernel within SVR needs to be positive semidefinite, while the positive semidefinite components can be either added or multiplied to generate effective kernel function for different real applications [7]. For the derived components, the solution to multiply them is adopted in this study to consider the correlation of both storage temperature and SoC terms as well. In this context, a coupled battery knowledge-guided kernel is finally created with the following specific form:

$$\begin{aligned} k_{\text{know}}(\boldsymbol{x}, \boldsymbol{x}') = & \, a \cdot \exp\left(b \cdot \|x_{Cap} - x_{Cap}'\|^2\right) \cdot \left(x_{\Delta t}^T x_{\Delta t}'\right)^z \\ & \cdot \exp\left(c \cdot \left\|\frac{1}{x_T} - \frac{1}{x_T'}\right\|\right) \cdot \left(d \cdot x_{SoC}^T x_{SoC}' + e\right) \end{aligned} \tag{11}$$

where a, b, c, d, e, and z are its related hyper-parameters, $\boldsymbol{x}$ represents input vector as $\boldsymbol{x} = \left(x_{Cap}, x_{\Delta t}, x_T, x_{SoC}\right)$. According to the data-driven model structure in Fig. 2, $x_{Cap} = Cap(t)$, $x_{\Delta t} = \Delta t_{fut}$, $x_T = T_{sto}$, $x_{SoC} = SoC_{sto}$.

After developing the coupled knowledge-guided kernel, the related SVR-based model can be trained by using a partial accelerated ageing dataset under high storage temperatures or SoCs that are easy to be collected. Then, it can be adopted directly for future calendar ageing trajectory prediction under the witnessed conditions that are similar as the training cases. For unwitnessed conditions where the storage temperature and SoC are different from training cases, to ensure good generalization of the developed data-driven model, a transfer part is involved.

Kernel for transfer part: To equip transfer concept into the SVR-based model for unwitnessed conditions, a transfer kernel is derived based on the Gaussian function as

$$k_{\text{transfer}}(\boldsymbol{x}, \boldsymbol{x}') = \lambda \cdot \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\sigma^2}\right) \tag{12}$$

where λ and σ are its hyper-parameters. For battery capacity trajectory prediction of the witnessed case, only $k_{\text{know}}(\boldsymbol{x}, \boldsymbol{x}')$ will be trained. For the prediction of the unwitnessed case, the well-trained $k_{\text{know}}(\boldsymbol{x}, \boldsymbol{x}')$ is fixed, while the $k_{\text{transfer}}(\boldsymbol{x}, \boldsymbol{x}')$ will be tuned by using only partial starting capacity data of the unwitnessed case as

$$k_{SVR}(\boldsymbol{x},\boldsymbol{x}') = \begin{cases} k_{\text{know}}(\boldsymbol{x},\boldsymbol{x}'), & \text{for witnessed case} \\ k_{\text{know}}(\boldsymbol{x},\boldsymbol{x}') + k_{\text{transfer}}(\boldsymbol{x},\boldsymbol{x}'), & \text{for unwitnessed case.} \end{cases} \tag{13}$$

Results and discussion: To evaluate the performance of derived model in battery calendar ageing trajectory predictions, case studies for both witnessed (same conditions as the training series) and unwitnessed conditions are carried out by using Matlab2022 with a 2.4 GHz Intel 4 CPU.

Case study for various temperatures: In this case study, accelerated ageing data including capacity points before 5000 hours of Cases 6–8 are utilized to train the data-driven model. Then the corresponding extrapolated prediction results after 5000 hours of these witnessed cases are shown in Fig. 3. The local increased battery SoH point is mainly caused by "capacity regeneration" and shifted measurement [8]. After using biogeography-based optimization (BBO) algorithm to minimize the prediction error, the optimized results of hyperparameters within the knowledge-guided kernel function are: $a = 0.0006$, $b = 0.7326$, $c = -0.7628$, $d = 0.0007$, $e = 0.0011$, $z = 0.46$. It can be noted that the predicted trajectories from the proposed model well match most real points of all these three cases with R^2 value over 0.98, indicating good ageing trajectory prediction performance for witnessed cases.

After using the related starting capacity data before 2000 hours to respectively tune the transfer part for all five unwitnessed cases (Cases 1−5), their corresponding battery future capacity trajectory prediction results are shown in Fig. 4. It can be seen that the developed model can also present good performance for such unwitnessed cases as all these five predicted trajectories can well match the real points in general with R^2 value over 0.97.

Comparisons with benchmarks: To further investigate the derived data-driven model's performance, another three classical models are utilized as the benchmarks for comparisons. Specifically, Benchmark 1 is an SVR model with the developed knowledge-guided kernel but without transfer concept, Benchmark 2 is an SVR model with just Gaussian-based kernel, while Benchmark 3 is a classical radial basis function (RBF)-based neural network. All models are trained and validated by following the same process of the above case study (capacity points before 5000 hours of Cases 6–8 are used for model training). Then Case 1 is used for model performance evaluation as it presents the worst results. Without the loss of generality, typical performance indicators including mean absolute error (MAE), root mean square error (RMSE), and R^2 are utilized to quantify the prediction results, as illustrated in Table 2.

Quantitatively, the proposed model with both battery knowledge-guided kernel and transfer concept achieves the best results among four models, whose R^2 is 2.1% better than that of Benchmark 1. This implies the effectiveness of transfer part to improve prediction performance for unwitnessed conditions. In contrast, without battery knowledge-guided kernel, both Benchmarks 2 and 3 present worse prediction performance with R^2 of 0.90 (7.2% decrease) and 0.89 (8.2% decrease), respectively. In light of this, the data-driven model coupling battery ageing knowledge can effectively improve the performance of battery calendar ageing trajectory prediction.

Conclusions: This paper presents a battery calendar ageing trajectory prediction strategy based on the knowledge-guided data-driven model with transfer concept. Illustrative results show that the proposed knowledge-guided data-driven model can provide accurate calendar ageing trajectory predictions for both witnessed and unwitnessed conditions. Due to the superiorities of being flexible, our future work plans to continue performing calendar ageing experiments to generate long-term battery ageing data from other battery types and varied temperature cases, while extending this knowledge-guided data-driven model with suitable feature terms for more battery ageing prognostic applications.

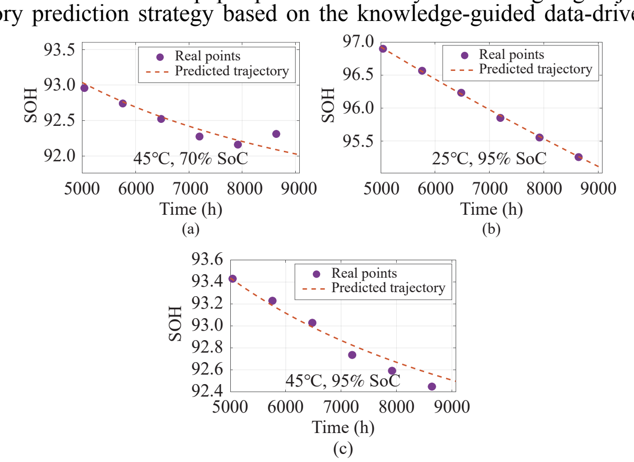

Fig. 3. Future ageing trajectory prediction results for witnessed cases: (a) Case 6; (b) Case 7; (c) Case 8.

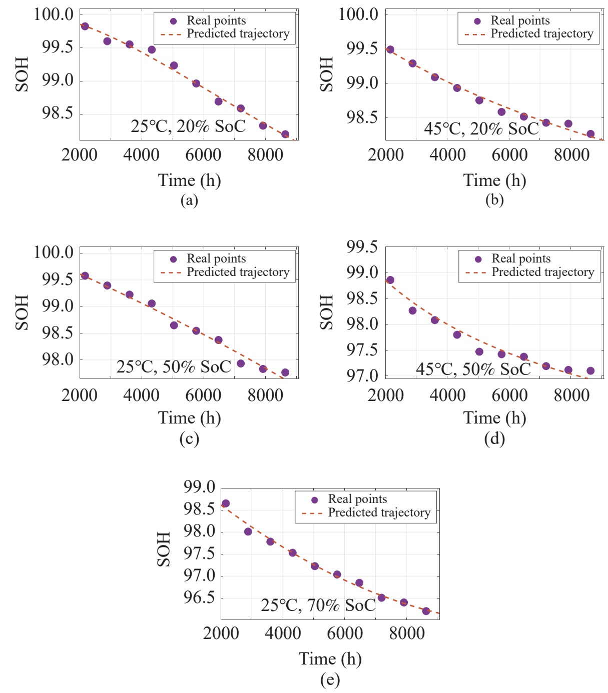

Fig. 4. Future ageing trajectory prediction results for unwitnessed cases: (a) Case 1; (b) Case 2; (c) Case 3; (d) Case 4; (e) Case 5.

Table 2. Comparison Results Under the Unwitnessed Condition

	MAE	RMSE	R^2
Proposed model	0.132	0.058	0.97
Benchmark 1	0.196	0.083	0.95
Benchmark 2	0.293	0.129	0.90
Benchmark 3	0.301	0.132	0.89

References

[1] K. Liu, Z. Wei, C. Zhang, Y. Shang, R. Teodorescu, and Q.-L. Han, "Towards long lifetime battery: AI-based manufacturing and management," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 7, pp. 1139–1165, Jul. 2022.

[2] K. Liu, *et al*., "Electrochemical modeling and parameterization towards control-oriented management of lithium-ion batteries," *Control Eng. Pract*, vol. 124, p. 105176, Apr. 2022.

[3] Y. Li, *et al*., "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," *Renew. Sust. Energ. Rev*, vol. 113, p. 109254, Jul. 2019. DOI: 10.1016/j.rser.2019.109254.

[4] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," *Joule*, vol. 4, no. 2, pp. 310−346, Feb. 2020.

[5] M. Dubarry, N. Qin, and P. Brooker. "Calendar aging of commercial Li-ion cells of different chemistries–A review." *Current Opinion Electrochemistry*, vol. 9, 106−113, 2018.

[6] T. Hu, H. Ma, H. Sun, and K. Liu, "Electrochemical-theory-guided modelling of the conditional generative adversarial network for battery calendar ageing forecast," *IEEE J. Emerg. Sel. Top. Power Electron*, Feb. 2022.

[7] B. Jiang, H. Dai, X. Wei, and Z. Jiang, "Multi-kernel relevance vector machine with parameter optimization for cycling aging prediction of lithium-ion batteries," *IEEE J. Emerg. Sel. Top. Power Electron*, Dec. 2021.

[8] K. Liu, Y. Shang, Q. Ouyang, and W. D. Widanage, "A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery," *IEEE Trans. Ind. Electron*, vol. 68, no. 4, pp. 3170–3180, Apr. 2021.