Human-knowledge-augmented Gaussian process regression for state-of-health prediction of lithium-ion batteries with charging curves

Zhou, Quan; Wang, Chongming; Sun, Zeyu; Li, Ji; Williams, Huw; Xu, Hongming

DOI:
10.1115/1.4050798

Document Version
Peer reviewed version

Citation for published version (Harvard):
Zhou, Q, Wang, C, Sun, Z, Li, J, Williams, H & Xu, H 2021, 'Human-knowledge-augmented Gaussian process regression for state-of-health prediction of lithium-ion batteries with charging curves', *Journal of Electrochemical Energy Conversion and Storage*, vol. 18, no. 3, 030907 . https://doi.org/10.1115/1.4050798

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

Human-knowledge-augmented Gaussian Process Regression for State-of-Health Prediction of Lithium-ion Batteries with Charging Curves

Quan Zhou
Research Fellow, Leader of CASE-V Research
Vehicle Research Centre
University of Birmingham
Birmingham, B15 2TT, UK
Email: q.zhou@bham.ac.uk

Chongming Wang
Assistant Professor
Institute for Future Transport and Cities
Coventry University
Coventry, CV1 5FB, UK
Email: ac8174@coventry.ac.uk

Zeyu Sun
PhD student
Vehicle Research Centre
University of Birmingham
Birmingham, B15 2TT, UK
Email: zxs994@bham.ac.uk

Ji Li
Research Fellow
Vehicle Research Centre
University of Birmingham
Birmingham, B15 2TT, UK
Email: j.li.1@bham.ac.uk

Huw Williams
Professor
Vehicle Research Centre
University of Birmingham
Birmingham, B15 2TT, UK
Email: h.williams.5@bham.ac.uk

Hongming Xu*
Professor and Head of Vehicle Research Centre
University of Birmingham
Birmingham, B15 2TT, UK
Email: h.m.xu@bham.ac.uk

Abstract

Lithium-ion batteries have been widely used in renewable energy storage and electric vehicles, and State-of-Health (SoH) prediction is critical for battery safety and reliability. Following the standard SoH prediction routine based on charging curves, a human-knowledge-augmented Gaussian process regression (HAGPR) model is proposed by incorporating two promising artificial intelligence techniques, i.e., the Gaussian process regression (GPR) and the adaptive neural fuzzy inference system (ANFIS). Human knowledge on voltage profile during battery degradation is firstly modelled with a ANFIS for feature extraction that helps reduce the need of physical testing. Then, the ANFIS is integrated with a GPR model to enable SoH prediction. Using a GPR model as the baseline, a comparison study is conducted to demonstrate the advantage of the proposed HAGPR model. It indicates that the proposed HAGPR model can reduce at least 12% root mean square error with 31.8% less battery aging testing compared to the GPR model.

Nomenclature

Abbreviations		RMSE	Root mean square errors
ABSE	Absolute errors	SoC	State of Charge
ANFIS	Adaptive neural fuzzy inference system	SoH	State of Health
BMS	Battery management system		
BTMS	Battery thermal management system	**Symbols**	
CC	Constant current	F	Feature
CV	Constant voltage	t	Time
EV	Electric vehicles	Δt	Sampling time
GPR	Gaussian process regression	M	Membership function

* Corresponding author: Professor Hongming Xu (h.m.xu@bham.ac.uk)

HAGPR	Human-knowledge augmented GPR	V	Voltage
HEV	Hybrid electric vehicle	$\mathcal{G}$	Gaussian process model
IEA	International energy agency	$\boldsymbol{v}$	Triangle parameter vector
OCV	Open circuit voltage	F_*	Extracted feature
PCC	Pearson correlation coefficient		

1. Introduction

The International Energy Agency (IEA) predicted that, in ten years' time, the share of renewables in global electricity supply will be 60% [1] and the global electrified vehicle stock will reach 245 million [2]. Lithium-ion batteries are core components that have been widely adopted in renewable energy storage and electrified transport systems, e.g., wearable equipment [3], [4], smart grids [5], [6], energy harvesting systems [7], [8], hybrid vehicles [9], [10], and electric vehicles [11], [12]. Estimation and prediction of battery states, e.g., State-of-charge (SoC) and state-of-health (SoH), are critical for design of battery management systems (BMS) [13], [14], battery thermal management systems (BTMS) [15], [16], and battery-integrated systems (e.g., hybrid vehicles [17], [18]) to allow safe and reliable operation of the lithium-ion batteries. It is also important for the control functionalities in energy systems, e.g., charge sustaining control [19], [20] and energy management system [21], [22] of electrified vehicles. Estimation and prediction of SoH is very important for battery aging judgment and can provide guidance for reasonable battery use [23], [24]. Compared to SoC, it is more challenge for estimation and prediction of SoH, where the former calculates the SoH at the current state and the later predicts the SoH in the future.

To enable accurate SoH estimation and prediction, analytic models and data-driven models have been developed as the two mainstreams in recent years. Analytic models, including physically-based electrochemical models, semi-empirical models, and empirical models are developed based on the physical and chemical mechanism of battery degradation [25]. Mevawalla et al. proposed a mathematical model based on experimental measurement of the tab/surface temperature, separator, electrolyte resistance, anode-cathode irreversible and reversible heat [26]. The analytic models have contributed developed understanding of battery degradation mechanisms, however, there are momentous challenges for these analytic models in simulation of all underlying dynamics in battery degradation with fast computing speed.

With rapid development in artificial intelligence (AI) and informatics technology (IT), it becomes feasible to estimate and predict SoH with fast simulation speed by using data-driven models which were mainly developed using parametric and non-parametric methods. The parametric methods are used for building support vector machine (SVM), relevance vector machine (RVM), and artificial neural networks (ANN). Nuhic et al. implemented a SVM for SoH estimation [27]. Hu et al. enabled online SoH estimation based on sparse Bayesian learning where a RVM is employed for probabilistic kernel regression [28]. Garg et al. proposed a genetic programming method for modelling of battery aging with multidisciplinary parameters [29]. Panchal implemented an ANN model for prediction of battery performance at different discharging rates and boundary temperature conditions [30].

With significant superior representation capability and prediction accuracy than parametric methods (e.g., SVM and ANN), Gaussian process regression (GPR) has been recognised as a promising machine learning technique in recent years [31]. As a well-known non-parametric method, GPR treats the input-to-output mapping as a random function with a probability density defined using a Gaussian process prior. Liu et al. used gaussian process regression to enable prognostics of battery SoH [32]. Li et al. implemented a mixed model of GPR and particle filtering to predict the SoH of a battery under uncertain conditions [33]. Richardson et al. proposed a GPR method for forecasting battery's SoH and highlighted various advantages of GPR over other data-driven and mechanistic approaches [34]. Yang et al. proposed a GPR model based on charging curves to enable battery SoH estimation [35]. Although GPR models can achieve promising results, obtaining robust AI-based SoH prediction models with less data and faster learning process is still a challenge task.

Feature extraction is an important technique for obtaining robust and accurate models. Conventionally, features are extracted by calculating the mean, maximum, and minimum values of the data in time domain or frequency domain. In recent years, feature extraction function has been integrated in the AI models. Tagade et al. developed a deep learning for battery capacity estimation, end of life prediction, and degradation mode diagnosis simultaneously, and a feature vector is introduced for feature extraction and trained as a part of the deep network [36]. In Tagade's model, the feature extraction is purely data-driven and in somehow increase the computational effort by introducing more parameters for neural network training. The authors developed a fuzzy feature extraction method for driving behavior identification [37]. By defining fuzzy rules and membership functions based on human knowledge, the fuzzy feature extraction method is shown superior to the conventional method with time domain and frequency domain data. Recently, adaptive neuro-fuzzy inference systems (ANFISs) have elaborated superiorities in machine learning as they can incorporate heuristic human knowledge in data-driven modelling [38]. As many theorical and experimental studies have been conducted, the

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

extractable features from human knowledge for battery state estimation and prediction can be the shape of charging voltage curves [35], chemical and physical mechanicians in battery degradation [25]. However, the research on feature extraction with the ANFIS for SoH prediction is scarcely reported.

To address the challenges as illustrated above, this paper proposed an improved GPR model that can incorporate human knowledge with a deep architecture. Following the standard battery diagnosis routine that is based charging curves [35], the present work is conducted with two original contributions: 1) An adaptive neural fuzzy inference system (ANFIS) is developed for the first time to enable feature extraction based on human's knowledge to reduce the need of physical battery aging testing; 2) A human-knowledge-augmented Gaussian process regression (HAGPR) model is proposed by incorporating the ANFIS with a Gaussian process model to enhance the SoH prediction capability.

The rest of this paper is organized as follows. Section 2 introduces the battery aging test setup, the testing datasets, and the features in the charging curves. The HAGPR network is proposed in Section 3. Section 4 discusses the results of the experimental evaluation on the proposed HAGPR network, and conclusions are summarized in Section 5.

2. Aging testing of lithium-ion batteries

2.1. Experiment setup and battery datasets

The cyclic aging datasets of lithium-ion batteries are obtained from an open-source data repository of the NASA Ames Prognostics Centre of Excellence (PoCE) [39]. The data was collected from a battery prognostics test bed at room temperature (24 ℃) with four commercially available 18650 battery cells, i.e., No.5, No.6, No.7, and No.18. The nominal voltage and capacity of the 18650 cell are 3.7 V and 2600 mAh, respectively [39]. The aging testing was conducted under several charge-discharge cycles. Each cycle started with a discharging process that was followed by a charging process. For the discharging process, the battery was discharged with a constant current (CC) until the battery voltage is less than the lower limit. The discharging current and voltage limit for different batteries are illustrated in Table.1 [40]. For the charging process, the battery was firstly charged with a CC of 1.5 A until the battery voltage reached 4.2 V and then charged with a constant voltage (CV) of 4.2 V until the battery current dropped to 30 mA.

Table.1. Discharge conditions of the studied batteries [39]

	Charging conditions		
	Voltage upper limit (V)	Voltage lower limit (V)	Discharge current (A)
Battery No.5	4.2	2.7	2
Battery No.6	4.2	2.5	2
Battery No.7	4.2	2.2	2
Battery No.18	4.2	2.2	4

The attenuation curves (SoH vs. number of cycles) of the studied four battery cells are illustrated in Fig.1. In general, battery SoH decreases during the aging testing because of battery degradation. Because of the regeneration phenome, the battery SoH is not a monotone decreasing form with the progressing of the testing cycles. This small-range capacity rise phenomenon has a significant influence on the accuracy and precision of SoH prediction. Therefore, investigations on the features in charging curves that influence SoH performance is necessary to obtain advanced SoH models.

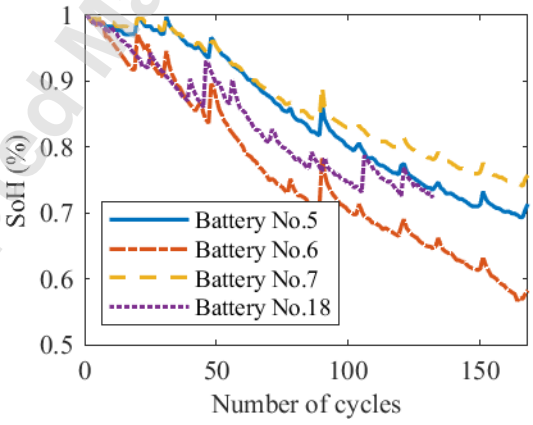

Fig. 1. Battery SoH degradation during the aging testing [39]

2.2. Representative features in charge curves

By looking into the aging testing data of battery No.6 as an example, the charging curves of cycle 1, 50, 80, and 130 are compared in Fig.2 (a), which indicate that the CC charging period has a decreasing trend while CV charging

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

period has an increasing trend as progressing of the aging testing. The graphical representations of the CC charging curves are also various during the aging testing. According to the geometrical analysis of charging curve in different number of cycles as in [35], four representative features, i.e., F1, F2, F3 and F4, as shown in Fig.2 (b), are selected for SOH prediction as defined in Table 2.

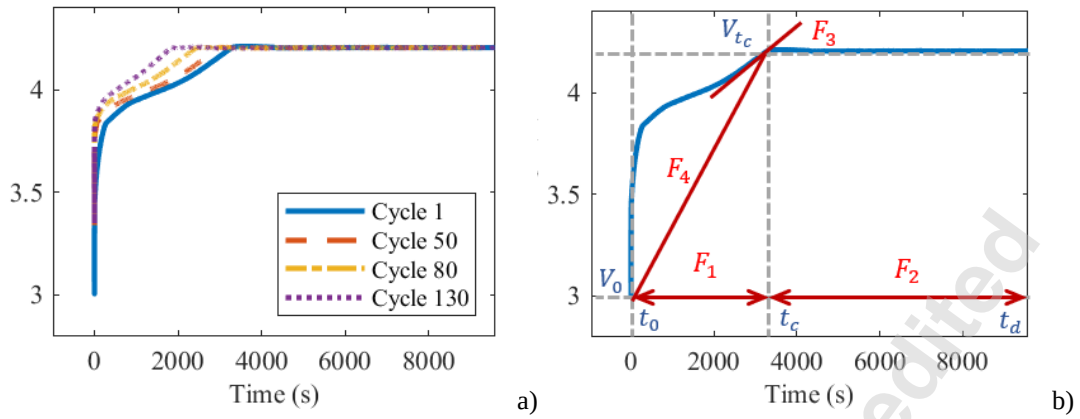

Fig. 2. Charging curves of battery No.6: a) charging curves at different cycles; b) representative features [35]

Table.2Definition of the four features in charging curves [35]

Feature	Description	Mathematical model
F_1	The time of CC charge stage	$F_1 = t_c - t_0$, where t_0 is the start time of the CCCV charging cycle; and t_c is the time when the battery voltage is firstly equal to the upper voltage limit of 4.2 V.
F_2	The time of CV charge stage	$F_2 = t_t - t_c$, where t_t is the time when the battery charging current is less than 30 mA.
F_3	The transient voltage changing rate at the boundary between CC stage and CV stage	$F_3 = \frac{V_{t_c} - V_{(t_c-1)}}{\Delta t}$, where Δt is the sampling time to obtain the charging curve; and $V_{(t_c-1)}$ is the measured voltage one sample step earlier than t_c.
F_4	The average voltage changing rate at the CC stage	$F_4 = \frac{V_{t_c} - V_0}{t_c - t_0}$, where V_0 is the battery voltage at the beginning of the CCCV cycle

3. Proposed methodology

Following the standard SoH estimation routine based on charging curves [35], the human-knowledge-augmented Gaussian process regression (HAGPR) method, as illustrated in Fig.3, is proposed with four main procedures, i.e., 1) preparation, 2) feature extraction, 3) SoH estimation, and 4) experimental evaluation, which are highlighted with blue, orange, green, and yellow, respectively.

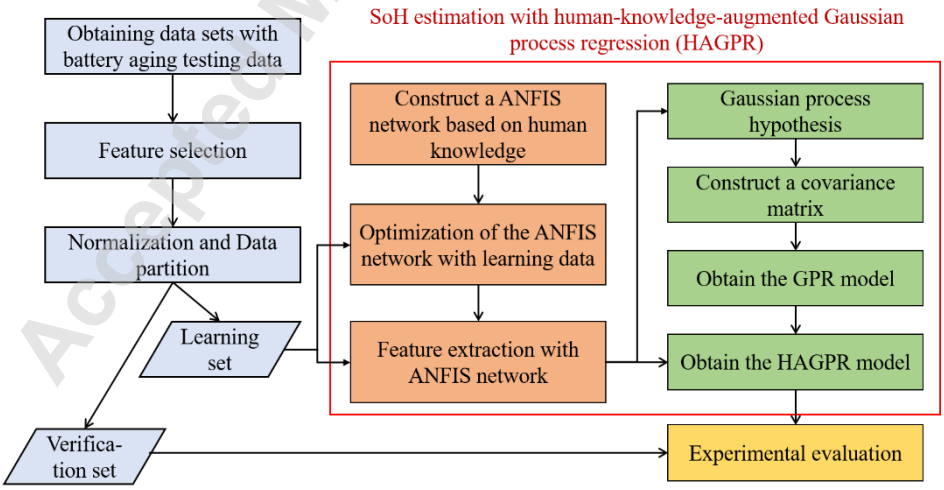

Fig. 3. Procedures of SoH estimation with the human-knowledge-augmented Gaussian process regression (HAGPR)

Firstly, the datasets obtained from battery aging testing are pre-processed and portioned into learning set and verification set. Four features, as identified in Section 2.2, are selected as the input of the SoH estimation model. The Min-Max method is used to normalize the feature values between 0 and 1 by

$$\boldsymbol{F}_f{}' = \frac{\boldsymbol{F}_f - \min(\boldsymbol{F}_f)}{\max(\boldsymbol{F}_f) - \min(\boldsymbol{F}_f)} \quad f = 1,2,3,4 \tag{1}$$

where, $\boldsymbol{F}_f$ is a feature vector of the fth feature ($f = 1,2,3,4$) that contains the feature values calculated with the battery aging test data using the equations in Section 2.2; $\min(\boldsymbol{F}_f)$ and $\max(\boldsymbol{F}_f)$ are the minimum value and the maximum value of the feature vector, respectively; and $\boldsymbol{F}_f{}'$ is the normalized feature vector.

Then, the proposed HAGPR model, as shown in Fig. 4, is developed by incorporating the ANFIS network (for feature extraction) and the GPR model (for SoH estimation). The inputs of the HAGPR model are the normalized four features, and the output is the battery SoH. The detailed procedures for ANFIS-based feature extraction and GPR-based SoH estimation are proposed as follows.

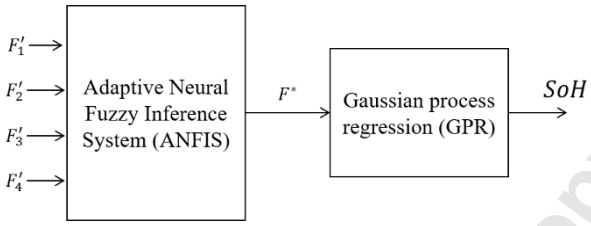

Fig. 4. Architecture of the HAGPR model

3.1. Feature extraction with adaptive neural fuzzy inference

The adaptive neural fuzzy inference (ANFIS) network is adopted for feature extraction, and a Takagi-Sugeno fuzzy model is used to build the ANFIS network, as illustrated in Fig.5, because it is easy to be implemented with data-driven learning [40]. The normalized feature values for each charging cycle are gathered in the input layer as an input vector $\boldsymbol{x} = [F_1', F_2', F_3', F_4']^T$, and the extracted feature, $F_* = y$, is generated in the output layer. The output, $y = \Phi(\boldsymbol{x})$, is calculated in three hidden layers using the input, $\boldsymbol{x}$. Three key components of a fuzzy inference system, i.e., fuzzification module, fuzzy rule base, and output inferencing module, are included in three hidden layers of the ANFIS network.

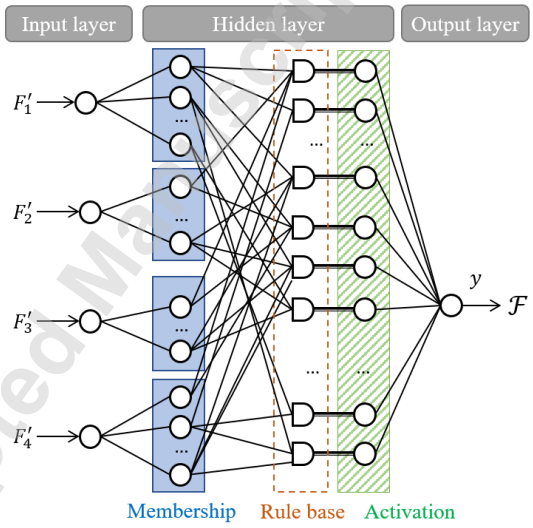

Fig. 5. Layout of the ANFIS network for feature extraction

The first hidden layer fuzzifies the inputs with triangular membership functions, $M_{1,h}$, $M_{2,i}$, $M_{3,j}$, and $M_{4,k}$, by

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

$$\left.\begin{aligned} M_{1,h}(F_1', \boldsymbol{v}_{1,h}) &= \max\left(\min\left(\frac{F_1'-\boldsymbol{v}_{1,i}(1)}{\boldsymbol{v}_{1,h}(2)-\boldsymbol{v}_{1,h}(1)} \quad \frac{\boldsymbol{v}_{1,h}(3)-F_1'}{\boldsymbol{v}_{1,h}(3)-\boldsymbol{v}_{1,h}(2)}\right), 0\right) \\ M_{2,i}(F_2', \boldsymbol{v}_{2,i}) &= \max\left(\min\left(\frac{F_2'-\boldsymbol{v}_{2,i}(1)}{\boldsymbol{v}_{2,i}(2)-\boldsymbol{v}_{2,i}(1)} \quad \frac{\boldsymbol{v}_{2,i}(3)-F_2'}{\boldsymbol{v}_{2,i}(3)-\boldsymbol{v}_{2,i}(2)}\right), 0\right) \\ M_{3,j}(F_3', \boldsymbol{v}_{3,j}) &= \max\left(\min\left(\frac{F_3'-\boldsymbol{v}_{3,j}(1)}{\boldsymbol{v}_{3,j}(2)-\boldsymbol{v}_{3,j}(1)} \quad \frac{\boldsymbol{v}_{3,j}(3)-F_3'}{\boldsymbol{v}_{3,j}(3)-\boldsymbol{v}_{3,j}(2)}\right), 0\right) \\ M_{4,k}(F_4', \boldsymbol{v}_{4,k}) &= \max\left(\min\left(\frac{F_4'-\boldsymbol{v}_{4,k}(1)}{\boldsymbol{v}_{4,k}(2)-\boldsymbol{v}_{4,k}(1)} \quad \frac{\boldsymbol{v}_{4,k}(3)-F_4'}{\boldsymbol{v}_{4,k}(3)-\boldsymbol{v}_{4,k}(2)}\right), 0\right) \\ h &= 1,2,\dots,H;\ i = 1,2,\dots,I;\ j = 1,2,\dots,J; k = 1,2,\dots,K; \end{aligned}\right\} \quad (2)$$

where, $M_{1,h}$ is the h-th membership function for the first feature; $M_{2,i}$ is the i-th membership function for the second feature; $M_{3,j}$ is the j-th membership function for the third feature; $M_{4,k}$ is the k-th membership function for the fourth feature; and $\boldsymbol{v}(l)$, l=1,2,3, is the l-th element of triangle parameter vectors, $\boldsymbol{v}$.

The second hidden layer connects the outputs of the input membership functions based on fuzzy rules. Each fuzzy rule applies the following linguistic logic

$$\begin{aligned} &\text{If } F_1' \text{ is } M_{1,h}(F_1', \boldsymbol{v}_{1,h}),\ F_2' \text{ is } M_{2,i}(F_2', \boldsymbol{v}_{2,i}),\ F_3' \text{ is } M_{3,j}(F_3', \boldsymbol{v}_{3,j}), \text{ and } F_4' \text{ is } M_{4,k}(F_4', \boldsymbol{v}_{4,k}), \\ &\quad \text{then } y \text{ is } L(\boldsymbol{x}, a_{h,i,j,k}), \end{aligned} \quad (3)$$

where $L(\boldsymbol{x}, a_{i,j})$ is the constant type membership function [37]; where $a_{i,j}$ is a scaling factor.

A vector of weighting values, $\boldsymbol{W} = [w_{1.1.1.1}, w_{1.1.1.2}, \dots, w_{1.1.1.K}, w_{1.1.2.1}, \dots, w_{H.I.1.1}, \dots, w_{H.I.J.K}]$, is used to infer the ANFIS output by weighting the outputs of the fuzzy rules:

$$y = \sum_{h=1}^{H}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left\{\min\left(M_{1,h}, M_{2,i}, M_{3,j}, M_{4,k}\right) \cdot L\left(\boldsymbol{x}, a_{h,i,j,k}\right) \cdot w_{h.i.j.k}\right\} \quad (4)$$

where $w_{h.i.j.k} \in [0,1], h = 1,2,\dots,H;\ i = 1,2,\dots,I;\ j = 1,2,\dots,J; k = 1,2,\dots,K;$

To build the ANFIS-based feature extraction model, $\mathcal{F} = \Phi([F_1', F_2', F_3', F_4']^T, \mathbb{C})$, a hyper parameter optimization is conducted to determine the parameter vector, $\mathbb{C} = [\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3, \boldsymbol{V}_4, \boldsymbol{A}, \boldsymbol{W}]$, where, $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $\boldsymbol{V}_3$ and $\boldsymbol{V}_4$ are parameter vectors of the inputs membership functions; $\boldsymbol{A}$ is the parameter vector of the output membership functions.

The hyper parameter optimization is based on the learning data set, $\{\boldsymbol{X} \quad Y\}$, where $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_N]$ and $Y = [y_1, y_2, y_3, .., y_N]$. To handle this hyper parameter optimization, a gradient free algorithm, i.e., generic algorithm, is used to determine the optimal parameter vector $\mathbb{C}^*$ that satisfies

$$\sqrt{\frac{\sum_{n=1}^{N}(\Phi(\boldsymbol{x}_n, \mathbb{C}^*) - y_n)^2}{N}} \leq \sqrt{\frac{\sum_{n=1}^{N}(\Phi(\boldsymbol{x}_n, \mathbb{C}) - y_n)^2}{N}} \quad (5)$$

where N is the number of data points in the dataset.

3.2. SoH estimation with Gaussian process regression

A Gaussian process model, $SoH = \mathcal{G}(\mathcal{F})$, is developed to estimate battery SoH with the extracted feature from the ANFIS network, $\mathcal{F} = \Phi([F_1', F_2', F_3', F_4']^T, \mathbb{C}^*)$. $\mathcal{G}(\boldsymbol{F}^*)$ is treated as a random function with a Gaussian prior, $\mathcal{gp}(\cdot)$, as

$$p\left(\mathcal{G}(\boldsymbol{F}^*)\right) = \mathcal{gp}(\mu(\boldsymbol{F}^*), K(\boldsymbol{F}^*, \boldsymbol{F}^*)) \quad (6)$$

where, $\boldsymbol{F}^*$ is a vector of all extracted feature values; $\mu(\boldsymbol{F}^*)$ is the mean function of $\boldsymbol{F}^*$; and $K(\boldsymbol{F}^*, \boldsymbol{F}^*)$ is the covariance function of $\boldsymbol{F}^*$. The GPR is trained by postulating a parametric form for the mean and the covariance function. The parameters of the mean and covariance functions are obtained as a hyper-parameter vector, $\boldsymbol{\theta}$, which is estimated by maximizing the log-likelihood of a training dataset $\{\boldsymbol{\mathcal{F}}^* \quad \mathbf{SoH}^*\}$.

During the learning process with the training dataset, predictions of $\mathbf{SoH}^*$ at a set of test inputs $\boldsymbol{\mathcal{F}}^*$ are obtained by computing a conditional probability distribution $p(\mathbf{SoH}^* | \mathbf{SoH}, \boldsymbol{F}^*, \boldsymbol{\mathcal{F}}^*, \theta)$, which is a multi-variate Gaussian, $\mathcal{N}(\tilde{\mu}, \tilde{K})$, with the mean function, $\tilde{\mu}$, and covariance function, $\tilde{K}$, given by

$$\left.\begin{aligned} \tilde{\mu}(\boldsymbol{\mathcal{F}}^*) &= \mu(\boldsymbol{\mathcal{F}}^*) + K(\boldsymbol{\mathcal{F}}^*, \boldsymbol{F}^*)K^{-1}(\boldsymbol{\mathcal{F}}^*, \boldsymbol{F}^*)[\mu(\boldsymbol{F}^*) - \mathbf{SoH}] \\ \tilde{K}(\boldsymbol{\mathcal{F}}^*, \boldsymbol{\mathcal{F}}^*) &= K(\boldsymbol{F}^*, \boldsymbol{F}^*) - K(\boldsymbol{\mathcal{F}}^*, \boldsymbol{F}^*)K^{-1}(\boldsymbol{F}^*, \boldsymbol{F}^*)K(\boldsymbol{F}^*, \boldsymbol{\mathcal{F}}^*) \end{aligned}\right\} \quad (7)$$

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

The computational effort in training the GPR model depends on the size of covariance matrix generated by $\widetilde{K}$. The most significant development in the present work is to reduce the covariance matrix size by reducing the number of features from four to one by implementation of the ANFIS network. As a data-driven learning, GPR is expected that the data points with similar feature values naturally have close SoH values. To reduce the negative impact of this similarity, kernel functions are widely adapted. This paper uses a squared exponential kernel function [31]

$$K(\boldsymbol{\mathcal{F}}^*, \boldsymbol{F}^*|\theta) = {\sigma_f}^2 \exp[-\frac{1}{2} \cdot \frac{(\boldsymbol{\mathcal{F}}^* - \boldsymbol{F}^*)^T(\boldsymbol{\mathcal{F}}^* - \boldsymbol{F}^*)}{{\sigma_l}^2}] \tag{8}$$

where $\theta = [\log \sigma_l, \log \sigma_f]$ is a parameter vector; σ_l is the characteristic length; and σ_f is the signal's standard derivation.

4. Experimental Evaluations

4.1. Person's correlation

By implementing five triangular membership functions for each of the input features, the ANFIS network for feature exaction is obtained based on the battery dataset No.6. Generic algorithm was used to determine the optimal parameter vector $\mathbb{C}^*$ that achieves the minimal root mean squared error (RMSE) between the ANFIS output and the experiment data. Six 3D maps were obtained in Fig.6.a) to Fig.6.f), respectively, which illustrate the inputs/output mapping of every two inputs to the exacted feature.

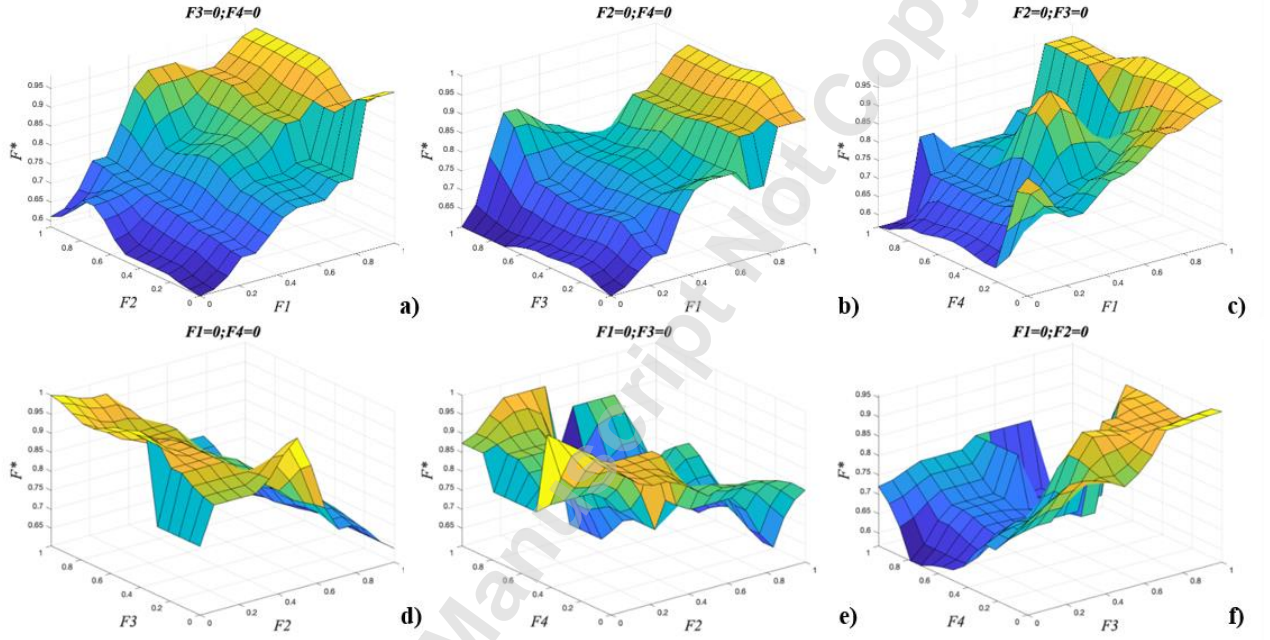

Fig. 6. Inputs/output mapping of the ANFIS: a) F1 and F2 to F*; b) F1 and F3 to F*; c) F1 and F4 to F*; d) F2 and F3 to F*; e) F2 and F4 to F*; f) F3 and F4 to F*

To study the feasibility of the proposed ANFIS-based feature extraction method, the ANFIS network obtained with dataset No.6 is used to extract features in datasets No.5, No.7, and No.18. Persons correlation coefficients (PCC) of the features (F_1', F_2', F_3', F_4', and F^*) and target (battery SoH) are obtained in the heatmaps as in Fig.6, where the PCC obtained with datasets No.6, No.5, No.7, and No.18 are illustrated in Fig.6.a), to Fig.6.d), respectively; and $F^*(N6)$ is the extracted feature based on the ANFIS model that is trained with the data set of battery No.6. The value (between 0 to 1) in the heatmap is the PCC value of its row labelled item and column labelled item, for example, the value of -0.321 in the second row and first column of Fig.5.a) is the PCC value of F_1' and F_2'. A PCC value closer to 1 means the two items linked to this CC value have stronger correlation.

The proposed ANFIS-based method is shown effective to obtain an extracted feature that has high correlation to the battery SoH (PCC value up to 0.9911). In most cases, the extracted feature has higher PCC value than the four original features. Conventional method in [35] needs all four features to estimate battery SoH because the PCC value rankings of these features to SoH are different in different datasets (e.g. F_4' has the second largest PCC value in dataset 6 but has the minimal PCC value in dataset 7) and some features have very similar PCC values (e.g. F_1' and F_3' in dataset 5). In the worst case (with dataset 18), the PCC value of the ANFIS extracted feature, $F^*(N6)$, (0.8881) is less than F_1' (0.909), the proposed method is still promising because it can reduce the feature number for SoH estimation from four to one.

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

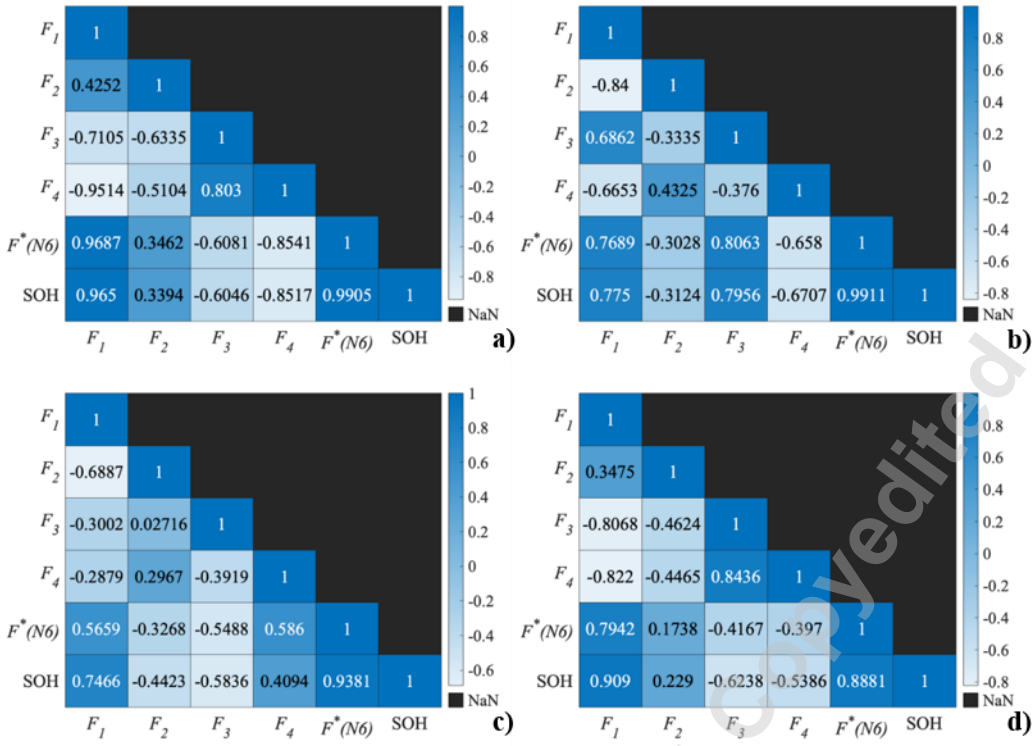

Fig. 7. Heatmaps of the Pearson correlation coefficients of the features and target: a) dataset No.6; b) dataset No.5; c) dataset No.7; and d) dataset No.18

4.2. Model learning performance

Based on the number of cycles, n_c, which indicates how many cycles the battery has been charged and discharged, the dataset is partitioned into two part for learning ($n_c = 1,2,\dots,N_c^l$) and validation ($n_c = N_c^l + 1, N_c^l + 2, \dots, N_c^t - 1, N_c^t$), where N_c^l is the number of cycles used for learning and N_c^t is the total number of cycles in an aging test. Learning performance of the proposed HAGPR model is evaluated by monitoring the estimated SoH (with the learning data), the predicted SoH (with the verification data), and the absolute errors with the dataset of battery No.6 for demonstration. The GPR model in [35] is chosen as baseline. With models obtained with different number of learning cycles (N_c^l =70, 75, 110, and 150), the learning performances are compared in Fig.8.a) to Fig.8.d), respectively.

Both conventional GPR and HAGPR have promising representative capability, and the absolute errors between the estimated SoH and real SoH in the learning space, $\mathrm{ABSE}_{\mathrm{est}} = |(\mathrm{SoH}_{\mathrm{est}} - \mathrm{SoH}_{\mathrm{real}})/\mathrm{SoH}_{\mathrm{real}}| \times 100\%$, can be control with more than 95% data located in the 5% error zone. The prediction performance depends highly on the numbers of cycle that are used for learning, and both of the conventional GPR model and the HAGPR model would have better prediction performance (fewer absolute errors with the validation data) if more cycle data is used for model learning. As shown by blue solid lines in Fig.8, the proposed HAGPR model has better prediction performance than the conventional GPR. This is because the human knowledge integrated in the ANFIS network can augment the information provided by the data, it can at least accurately predict the changing trend of the SoH curves when using little learning data, e.g., in Fig.7.a). With learning data more than in 75 testing cycles, the HAGPR model can achieve acceptable prediction results, i.e., the absolute prediction error can be control with more than 95% data in the 5% error zone.

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

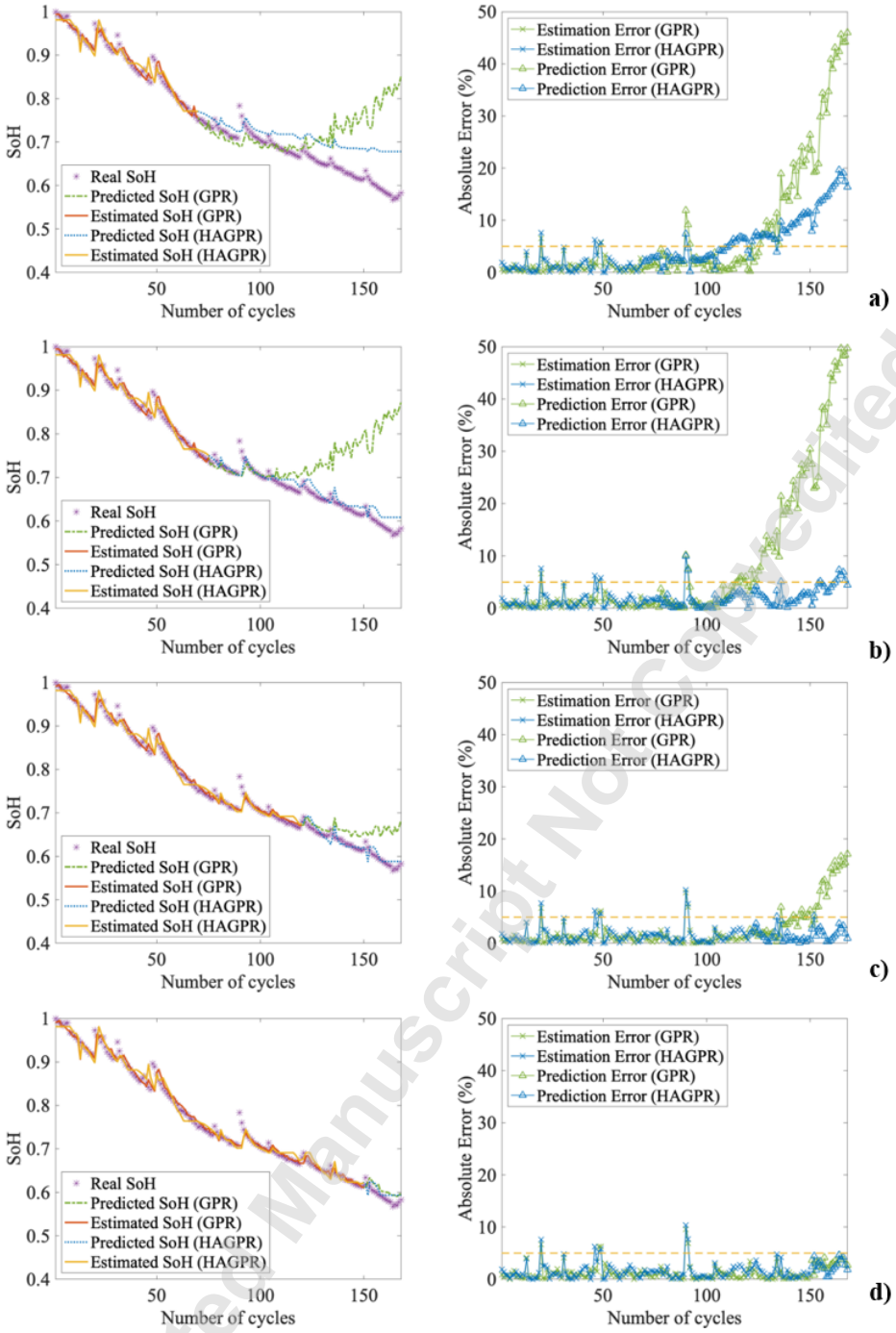

Fig. 8. Performance with different number of learning cycles: a) 70 cycles; b) 75 cycles; c) 110 cycles; and d)150 cycles.

The regression plots of GPR and HAGPR models for SoH estimation and prediction, which are obtained with different number of cycles for learning, are summarized in Fig.9. The coefficients of determination, i.e., the R squared values, are compared in Table 3. Both GPR and HAGPR have strong capability of SoH estimation by achieve very high R squired values (more than 0.9) when the inputs/output pairs are determined in the learning space for model training. In terms of SoH prediction, where the inputs/output data has not been used for model training, although both of them performs bad when the number of cycles for learning is 70, HAGPR is shown much better than conventional GPR by achieving higher R squared values when the number of cycles for learning is higher than 75.

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

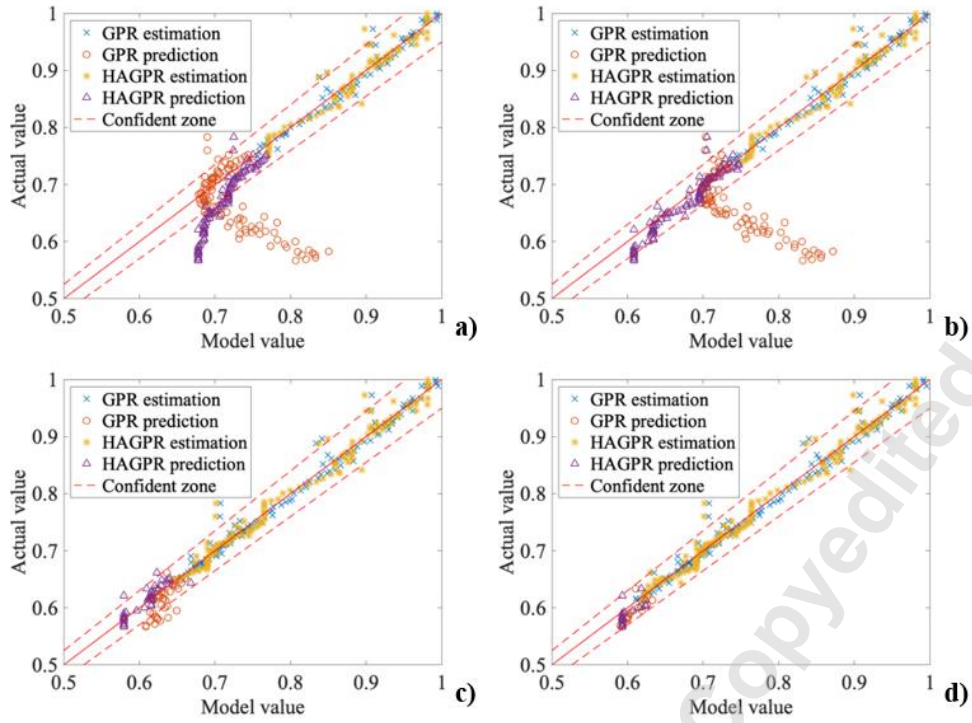

Fig. 9. Regression plots of different number of learning cycles: a) 70 cycles; b) 75 cycles; c) 110 cycles; and d)150 cycles.

Table.3 Coefficient of determination of the studied cases

Number of cycles for learning	Training		Validation	
	GPR	HAGPR	GPR	HAGPR
70	0.9485	0.9249	0.5800	0.8857
75	0.9598	0.9391	0.4417	0.9762
110	0.9718	0.9648	0.6840	0.9582
150	0.9836	0.9776	0.9849	0.9807

When using 'more than 95% predicted data is located in the 5% error zone' as the acceptance condition, the HAGPR model needs to conduct 75 cycles of aging testing to meet to acceptance condition while GPR needs 110 cycles. Therefore, the proposed HAGPR can save more than 31.8% aging testing to enable accurate battery SoH estimation compared to the conventional GPR method. In addition, this investigation shows that the HAPGR is capable of SoH prediction with a looking ahead window that has similar size to the number of cycles used for learning. This property of HAGPR can help determine the number of learning cycles for battery aging experiments, and k-fold cross-validation methods [7], [41] can be used to validate the SoH prediction model built with limited data.

4.3. Robustness of the SOH prediction for different batteries

To testify the robustness of the proposed HAGPR model in SoH prediction for different batteries, the estimated SoH, the predicted SoH, and the absolute errors are obtained with the datasets No.5, No.7, and No.8. To make the comparison (with the conventional GPR model) fair, both of the HAGPR model and GPR model are trained with the same data of 115 testing cycles. As illustrated in Fig.9, the proposed HAGPR can robustness outperformed the conventional GPR model by obtaining smoother and more accurate SoH predictions.

Journal of Electrochemical Energy Conversion and Storage. Received January 15, 2021;
Accepted manuscript posted March 30, 2021. doi:10.1115/1.4050798

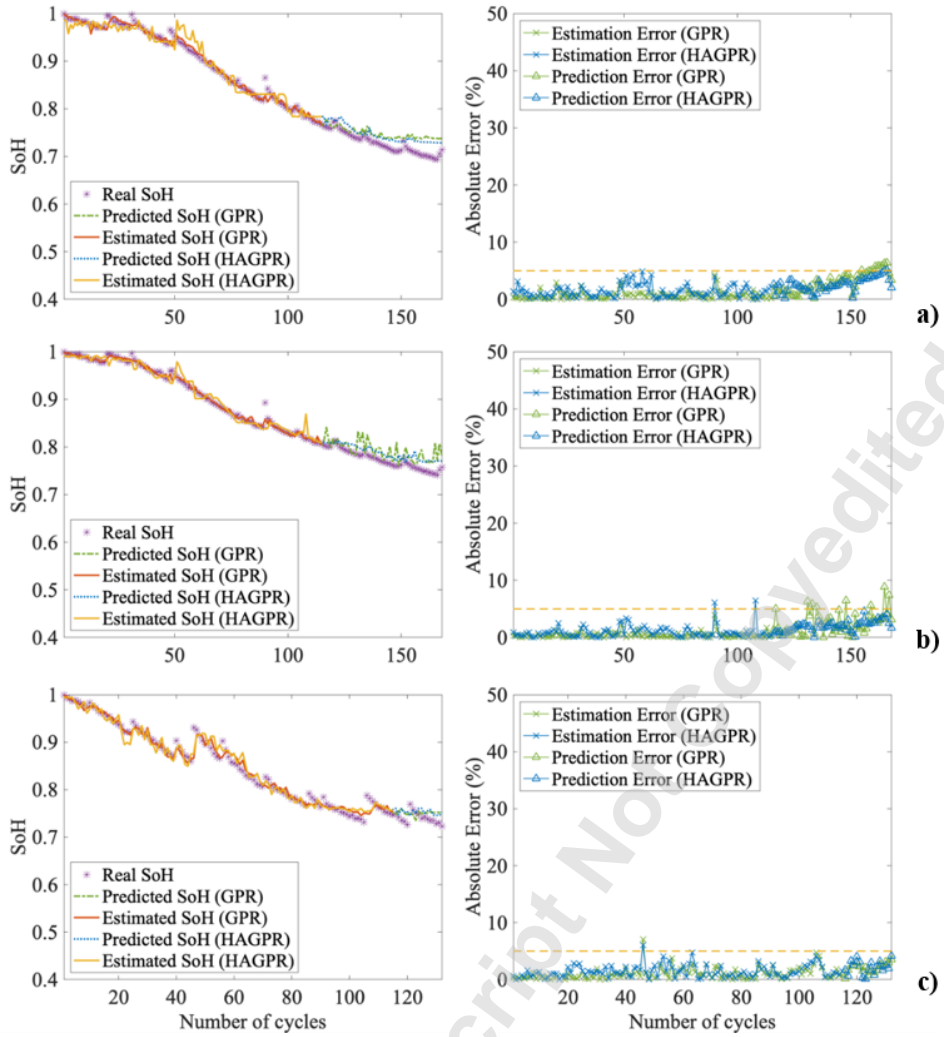

Fig. 10. Estimation and prediction performance with different batteries: a) No.5; b) No.7; and c) No.18.

By calculating mean squared errors (MSE) of the estimation results, $\mathrm{MSE_{est}} = (\mathrm{SoH_{est}} - \mathrm{SoH_{real}})^2/N_{est}$, and the prediction results, $\mathrm{MSE_{pre}} = (\mathrm{SoH_{pre}} - \mathrm{SoH_{real}})^2/N_{pre}$, where N_{est} and N_{pre} are the number of sample points for estimation and prediction, the SoH estimation and prediction performances for different batteries are quantified in Table.4. It indicated that the proposed HAGPR is shown more robust by achieving similar MSE values in SoH estimation and prediction. Although GPR can achieve lower MSE values in SoH estimation, the HAGPR model can robustly outperform the conventional GPR model for SoH prediction by reducing at least 25% MSE.

Table.4 Mean squared errors of estimated SoH and predicted SoH

	$\mathrm{MSE_{est}}$		$\mathrm{MSE_{pre}}$	
	GPR	HAGPR	GPR	HAGPR
Battery No.5	0.08e-3	0.24e-3	0.94e-3	0.25e-3
Battery No.7	0.03e-3	0.15e-3	2.9e-3	0.9e-3
Battery No.18	0.16e-3	0.23e-3	0.4e-3	0.3e-3

5. Conclusions

This paper proposed a human-knowledge-augmented Gaussian process regression (HAGPR) model, which adopts a ANFIS network to involve human knowledge on battery degradation to enable feature extraction for battery SoH prediction with Gaussian process regression. Based on the aging testing data of four selected Lithium-ion batteries, a comparison study is conducted to demonstrate the advantage of the proposed HAGPR model over a conventional GPR model in battery SoH prediction. The conclusions drawn from this research can be summarized as follows:

1) By introducing the ANFIS network based on the Takagi-Sugeno model, the HAGPR is capable of modelling of the Gaussian process in battery aging with the extracted feature. Pearson correlation tests have suggested that the feature extracted by the ANFIS network has improved correlations with SoH (the PCC value can be up to 0.9911).
2) The proposed HAGPR has the capability to enable accurate SoH prediction with less learning data than the conventional GPR model. According to the performance evaluation on the models developed with different number of cycles for learning, HAPGR can save more than 31.8% aging testing compared to the GPR model.
3) The proposed HAGPR model can robustly outperform the conventional GPR model for SoH prediction. At least 25% MSE can be reduced for SoH prediction of the selected Lithium-ion batteries.

In the planned future work, the HAGPR model will be extended for online SoH prediction with real-time battery data in vehicle daily use. The upgraded SoH model can contribute to multiple objective optimization and real-time energy management control of electrified vehicles.

Reference

[1] IEA, *World Energy Outlook 2020*. OECD, 2020.
[2] IEA, *Global EV Outlook 2020*. OECD, 2020.
[3] T. Ghomian and S. Mehraeen, "Survey of energy scavenging for wearable and implantable devices," *Energy*, vol. 178, pp. 33–49, 2019.
[4] J. Mi, Q. Li, M. Liu, X. Li, and L. Zuo, "Design, modelling, and testing of a vibration energy harvester using a novel half-wave mechanical rectification," *Appl. Energy*, vol. 279, no. July, p. 115726, 2020.
[5] L. Zhang *et al.*, "Hybrid electrochemical energy storage systems: An overview for smart grid and electrified vehicle applications," *Renew. Sustain. Energy Rev.*, vol. 139, no. October 2020, p. 110581, 2020.
[6] P. Lissa, C. Deane, M. Schukat, F. Seri, M. Keane, and E. Barrett, "Deep reinforcement learning for home energy management system control," *Energy AI*, vol. 3, p. 100043, 2021.
[7] Q. Zhou *et al.*, "Global Optimization of the Hydraulic-electromagnetic Energy-harvesting Shock Absorber for Road Vehicles with Human-knowledge-integrated Particle Swarm Optimization Scheme," *IEEE/ASME Trans. Mechatronics*, vol. 4435, no. c, pp. 1–1, 2021.
[8] T. Y. Kim, J. Kwak, and B. wook Kim, "Application of compact thermoelectric generator to hybrid electric vehicle engine operating under real vehicle operating conditions," *Energy Convers. Manag.*, vol. 201, no. October, p. 112150, 2019.
[9] Q. Zhou *et al.*, "Modified Particle Swarm Optimization with Chaotic Attraction Strategy for Modular Design of Hybrid Powertrains," *IEEE Trans. Transp. Electrif.*, vol. Early acce, 2020.
[10] Y. Huang *et al.*, "A review of power management strategies and component sizing methods for hybrid vehicles," *Renew. Sustain. Energy Rev.*, vol. 96, no. April 2017, pp. 132–144, 2018.
[11] C. Lv, X. Hu, A. Sangiovanni-Vincentelli, Y. Li, C. M. Martinez, and D. Cao, "Driving-Style-Based Codesign Optimization of an Automated Electric Vehicle: A Cyber-Physical System Approach," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 2965–2975, 2019.
[12] Y. Xing and C. Lv, "Dynamic State Estimation for the Advanced Brake System of Electric Vehicles by using Deep Recurrent Neural Networks," *IEEE Trans. Ind. Electron.*, vol. PP, no. c, p. 1, 2019.
[13] O. Teichert, F. Chang, A. Ongel, and M. Lienkamp, "Joint optimization of Vehicle Battery Pack Capacity & Charging Infrastructure for Electrified Public Bus Systems," *IEEE Trans. Transp. Electrif.*, vol. 5, no. 3, pp. 1–1, 2019.
[14] D. Sun *et al.*, "State of charge estimation for lithium-ion battery based on an Intelligent Adaptive Extended Kalman Filter with improved noise estimator," *Energy*, vol. 214, p. 119025, 2021.
[15] V. G. Choudhari, A. S. Dhoble, and S. Panchal, "Numerical analysis of different fin structures in phase change material module for battery thermal management system and its optimization," *Int. J. Heat Mass Transf.*, vol. 163, p. 120434, 2020.
[16] R. Jilte, A. Afzal, and S. Panchal, "A novel battery thermal management system using nano-enhanced phase change materials," *Energy*, vol. 219, p. 119564, 2021.
[17] M.-K. Tran, M. Akinsanya, S. Panchal, R. Fraser, and M. Fowler, "Design of a Hybrid Electric Vehicle Powertrain for Performance Optimization Considering Various Powertrain Components and Configurations," *Vehicles*, vol. 3, no. 1, pp. 20–32, 2020.
[18] Q. Zhou, W. Zhang, S. Cash, O. Olatunbosun, H. Xu, and G. Lu, "Intelligent sizing of a series hybrid electric power-train system based on Chaos-enhanced accelerated particle swarm optimization," *Appl. Energy*, vol. 189, pp. 588–601, 2017.
[19] B. Shuai *et al.*, "Heuristic action execution for energy efficient charge-sustaining control of connected hybrid vehicles with model-free double Q-learning," *Appl. Energy*, vol. 267, 2020.
[20] Q. Zhou, Y. Zhang, Z. Li, J. Li, H. Xu, and O. Olatunbosun, "Cyber-Physical Energy-Saving Control for Hybrid Aircraft-Towing Tractor based on Online Swarm Intelligent Programming," *IEEE Trans. Ind. Informatics*, vol. 14, no. 9, pp. 4149–4158, 2018.

[21] Y. Zhou, A. Ravey, and M. Péra, "A survey on driving prediction techniques for predictive energymanagement of plug-in hybrid electric vehicles," *J. Power Sources*, vol. 412, no. October 2018, pp. 480–495, 2019.
[22] Y. He *et al.*, "Multiobjective Co-Optimization of Cooperative Adaptive Cruise Control and Energy Management Strategy for PHEVs," *IEEE Trans. Transp. Electrif.*, vol. 6, no. 1, pp. 346–355, 2020.
[23] C. Pastor-Fernández, T. F. Yu, W. D. Widanage, and J. Marco, "Critical review of non-invasive diagnosis techniques for quantification of degradation modes in lithium-ion batteries," *Renew. Sustain. Energy Rev.*, vol. 109, no. April, pp. 138–159, 2019.
[24] R. Xiong, Y. Pan, W. Shen, H. Li, and F. Sun, "Lithium-ion battery aging mechanisms and diagnosis method for automotive applications: Recent advances and perspectives," *Renew. Sustain. Energy Rev.*, vol. 131, no. 5, p. 110048, 2020.
[25] C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, "Degradation diagnostics for lithium ion cells," *J. Power Sources*, vol. 341, pp. 373–386, 2017.
[26] A. Mevawalla, S. Panchal, M. K. Tran, M. Fowler, and R. Fraser, "Mathematical heat transfer modeling and experimental validation of lithium-ion battery considering: Tab and surface temperature, separator, electrolyte resistance, anode-cathode irreversible and reversible heat," *Batteries*, vol. 6, no. 4, pp. 1–26, 2020.
[27] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, "Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods," *J. Power Sources*, vol. 239, pp. 680–688, Oct. 2013.
[28] C. Hu, G. Jain, C. Schmidt, C. Strief, and M. Sullivan, "Online estimation of lithium-ion battery capacity using sparse Bayesian learning," *J. Power Sources*, vol. 289, pp. 105–113, 2015.
[29] A. Garg, S. Shaosen, L. Gao, X. Peng, and P. Baredar, "Aging model development based on multidisciplinary parameters for lithium-ion batteries," *Int. J. Energy Res.*, vol. 44, no. 4, pp. 2801–2818, 2020.
[30] S. Panchal, I. Dincer, M. Agelin-Chaab, R. Fraser, and M. Fowler, "Experimental and theoretical investigations of heat generation rates for a water cooled LiFePO4 battery," *Int. J. Heat Mass Transf.*, vol. 101, pp. 1093–1102, 2016.
[31] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
[32] D. Liu, J. Pang, J. Zhou, Y. Peng, and M. Pecht, "Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression," *Microelectron. Reliab.*, vol. 53, no. 6, pp. 832–839, 2013.
[33] F. Li and J. Xu, "A new prognostics method for state of health estimation of lithium-ion batteries based on a mixture of Gaussian process models and particle filter," *Microelectron. Reliab.*, vol. 55, no. 7, pp. 1035–1045, 2015.
[34] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Gaussian process regression for forecasting battery state of health," *J. Power Sources*, vol. 357, pp. 209–219, 2017.
[35] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *J. Power Sources*, vol. 384, no. September 2017, pp. 387–395, 2018.
[36] P. Tagade *et al.*, "Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis," *J. Power Sources*, vol. 445, no. October 2019, p. 227281, 2020.
[37] J. Li, Q. Zhou, Y. He, H. Williams, and H. Xu, "Driver-identified Supervisory Control System of Hybrid Electric Vehicles based on Spectrum-guided Fuzzy Feature Extraction," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 11, pp. 2691–2701, 2020.
[38] P. M. Ashok Kumar and V. Vaidehi, "A transfer learning framework for traffic video using neuro-fuzzy approach," *Sadhana - Acad. Proc. Eng. Sci.*, vol. 42, no. 9, pp. 1431–1442, 2017.
[39] B. Saha and K. Goebel, "Battery Data Set," NASA Ames Prognostics Data Repository, Moffett Field, CA, 2007.
[40] J. Cervantes, W. Yu, S. Salazar, and I. Chairez, "Takagi-Sugeno Dynamic Neuro-Fuzzy Controller of Uncertain Nonlinear Systems," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1601–1615, 2017.
[41] C. Lv *et al.*, "Hybrid-Learning-Based Classification and Quantitative Inference of Driver Braking Intensity of an Electrified Vehicle," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5718–5729, 2018.