

Useful energy prediction model of a Lithium-ion cell operating on various duty cycles

Damian Burzyński^a

^aPoznan University of Technology, Faculty of Control, Robotics and Electrical Engineering, ul. Piotrowo 3a, 60-965 Poznan, Poland

Indexed by:




Highlights

- A new non-parametric useful energy model for long-term prediction was developed.
- Developed model takes into account the lifetime degradation of the cell.
- Identification of three types of RUE_c evolution over exploitation period of the cells.
- XAI techniques were used to quantify effect of model parameters on RUE_c .
- The proposed methodology can be applied to electrochemical cells of other types.

Abstract

The paper deals with the subject of the prediction of useful energy during the cycling of a lithium-ion cell (LIC), using machine learning-based techniques. It was demonstrated that depending on the combination of cycling parameters, the useful energy (RUE_c) that can be transferred during a full cycle is variable, and also three different types of evolution of changes in RUE_c were identified. The paper presents a new non-parametric RUE_c prediction model based on Gaussian process regression. It was proven that the proposed methodology enables the RUE_c prediction for LICs discharged, above the depth of discharge, at a level of 70% with an acceptable error, which is confirmed for new load profiles. Furthermore, techniques associated with explainable artificial intelligence were applied to determine the significance of model input parameters – the variable importance method – and to determine the quantitative effect of individual model parameters (their reciprocal interaction) on RUE_c – the accumulated local effects model of the first and second order.

Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>) 

cycle life modelling, lithium-ion battery, machine learning, predictive models, useful energy prediction.

1. Introduction

In recent years, it is possible to observe a rapid increase in the demand for equipment designed for electric energy storage. At present, it is estimated that the highest increase refers to the industry branch related to electromobility, where solutions based on lithium-ion batteries are currently dominant [40]. It is predicted that by the end of 2030, the demand for batteries will increase ten times in this sector alone [13, 41].

The prevailing share of lithium-ion batteries on the market follows primarily from their ability to transfer energy quickly and effectively, whereby this ability is successively increased owing to the dynamic development of electrochemical battery technology [7].

The key issue related to the operation of these storages is their life. Degradation of lithium-ion cells (LICs) is a consequence of aging processes which take place during cycling, and also during the storage period [4, 54]. Out of more than a dozen identified aging phenomena, lithium plating, formation, evolution and dissolution of the solid electrolyte interface, as well as electrolyte decomposition, particle gassing and corrosion of the current collectors are the pre-dominant contributors to the degradation process [56]. Without interference in the LIC structure, the effects of the aging processes may be determined through the loss in charge throughput and the increase in internal resistance (or impedance).

1.1. Related papers

Based on the current state of knowledge, it can be concluded that in the case of LICs, usually separate degradation models are developed for cycling (cycle life) and storage period (calendar aging). The calendar aging models usually include two parameters: the ambient temperature (T_a) at which a cell is stored and the state of charge (SoC) [2, 3, 23, 64]. On the other hand, degradation models during cycling are multi-parameter models – their parameters include: values of discharge/charge current (I_d/I_{ch}), depth of discharge (DoD), cell temperature (T_c) or ambient temperature, as well as the charge or energy throughput (expressed most frequently as the number of full equivalent cycles – FECs). At this point, it must be emphasised that some of these parameters are correlated with each other and their effect on the amount of energy throughput of LICs is non-linear [38]. The above-mentioned facts make analysing and predicting the LICs' ability to transfer energy over the period of their lifetime highly complicated and require advanced methods [21]. In addition to this, testing electrochemical cells is very time-consuming due to their life. The testing of a single variant may last even a few years. For this reason, a more and more frequent practice is to conduct the testing under conditions which result in their accelerated degradation (e.g. under an increased load or at extreme temperatures), and then to predict their operational parameters under the assumed operating conditions [8, 12].

E-mail addresses: D. Burzyński (ORCID: 0000-0001-7874-1376); damian.burzynski@put.poznan.pl

It is possible to distinguish two main approaches in the modelling of LICs' degradation: physico-chemical models and empirical models. The physico-chemical models usually consist of several partial differential equations, by means of which, the effect of the respective aging processes on the capacity fade of LICs is modelled. Examples of such models, for the Li_xCoO_2 cell can be found in papers [37, 45, 50]. The advantage of physico-chemical models is the possibility of mapping changes taking place in each part of a LIC as a consequence of cycling, however, the identification of their parameters and implementation are highly complicated. A frequent approach among researchers is the use of methods based on regression [16, 36, 48, 53, 58]. In such models, in the majority of cases, the effect of selected cycling parameters on LIC degradation is analysed. A paper by [15] studied the effect of certain parameters of the charging half-cycle, the charge currents (I_{ch}) and final charge voltages (U_{ch}), on the process of the actual capacity fade of a LiCoO_2 cell. The testing was conducted at one charge current level (1C) and at a temperature of 25°C. Xiong *et al.* [67] developed a model which enables prediction of the remaining useful life of a battery (RUL) for a LiNiCoAlO_2 cell. The tests were conducted under conditions of accelerated aging – at discharge currents $I_d = 1\text{C}$ and $I_d = 2\text{C}$ and at temperatures $T_a = 25^\circ\text{C}$ and $T_a = 40^\circ\text{C}$. Effect of fluctuating ambient temperature on the capacity loss process of LiFePO_4 cell was investigated and modeled in the paper [33]. In paper [35] Dessaint *et al.* implemented two multi-parametric models which allow for the determination of the maximum number of cycles which may be carried out by LiNiMnCoO_2 and LiFePO_4 cells, depending on the actual cycling parameters (I_d, I_{ch}, DoD, T_a). On the other hand, in paper [57] Wang *et al.* took into account the charge throughput in addition to typical cycling parameters (I_d, I_{ch}, DoD, T_c) to determine the capacity fade of a LiFePO_4 cell.

Among other approaches related to the modelling of degradation during LIC cycling, it is necessary to mention the destructive method based on the Palmgren-Miner damage accumulation theory [11,46] and the non-destructive method based on electrochemical impedance spectroscopy (EIS). Zhang *et al.* [69] proposed a forecasting methodology for the RUL of lithium-ion batteries using EIS and Gaussian process regression (GPR). In turn, Saha *et al.* in [47] used EIS and the Bayesian statistical approach to develop a new RUL method.

Moreover, as methods based on machine intelligence are being developed, it is possible to observe their more and more frequent use in problems related to the prediction of operational parameters of LICs such as, e.g. remaining useful life (RUL) [17, 32, 51] or capacity [43, 49]. These methods are strictly based on data which are usually acquired from many-months of experimental procedures [5, 10, 22, 28, 55]. In [19], Hannan *et al.* proposed the use of a deep neural network to determine the *SoC* for LiNiMnCoO_2 cells. Li *et al.* [29] developed a method for estimating the capacity of LiFePO_4 cells, based on a convolutional neural network. Moreover, the use of support vector machines was proposed in papers by [26, 63] to predict the state of health (*SoH*) of batteries. In [24], Li *et al.* proved that the *SoC* of a battery may be predicted with high accuracy by a structure which is a merger of the neural network and fuzzy logic. In the past, approaches based on sample entropy [20], fuzzy logic [6] and Rao-Blackwellization particle filter [14] have also been used to determine the RUL.

The GPR technique used in that paper had previously been used to predict the capacity fade or RUL. In [44], Howey *et al.* developed a model to predict the capacity fade of a LiCoO_2 cell under variable load and temperature conditions. On the other hand in [52] Hariharan *et al.* used the deep GPR to estimate the end of life (*EoL*) – the chemical composition of the cell was not specified. GPR was also used to predict battery degradation during calendar aging, and this issue was touched upon in paper by [31].

Because of the fact that, for some designers of systems powered from cells, *SoH* is not always a sufficient assessment indicator, other methods which enable the determination and prediction of the avail-

able power and energy in LICs have been developed recently [39, 65]. In [66], an attempt was made to predict the energy available in the battery for two cells - LiNiMnCoO_2 and LiFePO_4 , using particle filtering. In a paper by [25], the state of the available energy at different I_d and at different T_a , taking into account degradation, was determined for a $\text{Li}_4\text{Ti}_5\text{O}_{12}$ cell using a circuit model. Dong *et al.* [9] developed a state of energy estimation method for LiFePO_4 cells using neural network achieving an error of less than 4%. The analyses were conducted at different ambient temperatures and discharge currents. In the papers [30,68] methods based on predictive control theory were implemented to predict the energy of an LiNiMnCoO_2 cell and a package of LiNiMnCoO_2 cells. Another approach using an adaptive method based on an extended Kalman filter to estimate the remaining energy of LiFePO_4 and LiNiMnCoO_2 cells are presented in [60, 61]. Wang *et al.* [59] proposed a joint estimator based on particle filter to determine both the state of energy and *SoC* of the LiFePO_4 cell. It should be noted here that the cell parameters studies were conducted with variation in discharge currents and at four ambient temperature levels. Fractional-order physics models can be also used to determine the state of energy with less than 5% error, as demonstrated in the paper [27].

1.2. Key contributions

Although many different approaches have been developed to date in the aspect of modelling the effects of aging phenomena occurring in LICs, in particular the prediction of available energy and power, this issue has not yet been sufficiently explored. In many applications, not only is knowledge of the predicted useful energy of the LIC over its lifetime required, but also of the predicted changes in degradation trends. In many research centres around the world, complex models are currently being developed that take into account many parameters of cyclic operation, such as ambient temperature, discharge current and, increasingly *SoH*. For this reason, the following contributions have been made in this paper:

- 1) A major contribution is the development of a machine learning based model that allows long-term prediction of the useful energy that the LIC is able to transfer during a single duty cycle taking into account its lifetime degradation which is a significant improvement of currently existing methods. The proposed model has a new structure of input parameters – given the negative effect of higher currents during charging proven in the literature [15], the model separately considers the effect of discharging and charging current.
- 2) The result of this research is a new dataset that can be used to test current methods and develop new predictive methods for the useful energy or *SoH* of LICs.
- 3) A demonstration of the variability of RUE_c depending on the combination of cycling parameters, and also identification of three types of evolution of changes in RUE_c in the period of operation of the LIC under consideration.
- 4) For the first time, the following techniques have been used: explainable artificial intelligence (XAI) – variable importance (VI) – to determine the significance of parameters of the GPR model, and accumulated local effects (ALE) – to determine their quantitative effect on RUE_c .

1.3. Structure of the paper

The remaining part of the paper is organised as follows. Chapter II outlines the procedure for obtaining experimental data and determining the useful energy RUE_c and also describes the results of preliminary analyses and calculations. Chapter III presents the applied methodology, GPR and XAI techniques. The results obtained and the discussion of the verification procedure are included in chapter IV. The final remarks and the conclusion are presented in chapter V.

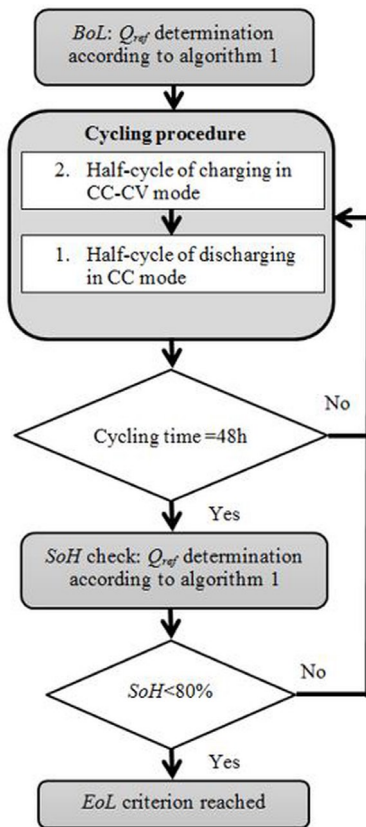


Fig. 1. Block diagram of the experimental procedure

Algorithm 1: Q_{ref} determination

1. Interruption of the currently running aging variant after completion of the charge half-cycle (not applicable to BoL)
2. Set the temperature of the cell to 25°C
3. Charging of the cell to 4.2 V with 1.3 A current in accordance with procedure CC-CV until the current in the CV phase drops to 0.065 A
4. 15-second break
5. Discharge of the cell to 2.75 V with the -2.6 A current (CC method)
6. Calculation of the cell capacity using the Coulomb counting method – equation (2)

2. Data preparation

In order to obtain data related to the ability of LICs to transfer energy during cycling, they were subjected to aging tests under various load and temperature conditions. Commercial Samsung 18650 LICs (cylindrical), with a nominal capacity of 2600 mAh (C_{nom}) and nominal voltage of 3.63 V (U_{nom}) were selected for the tests. The cathode material of the tested cells was composed of $\text{LiNi}_{0.33}\text{Mn}_{0.33}\text{Co}_{0.33}\text{O}_2$ compound, while the anode was made of graphite.

2.1. Experimental procedure

The procedure presented in Fig. 1 was adopted for the purpose of performing aging tests. Before starting the aging procedure of each cell, their reference tests charge (Q_{ref}) was determined in accordance with algorithm 1.

Then, each of the cells was subjected to cycling under unique temperature and load conditions. During the tests, a Panasonic MIR-254 temperature chamber was used to maintain a constant ambient temperature. The T_a range in which the cells were cycled was between 10°C and 40°C. Momentary temperatures of cells were recorded by means of probes located in central points of the cells. The discharge half-cycles were completed using the constant-current (CC) method, while charge half-cycles were completed using the constant-current constant-voltage (CC-CV) method. Val-

ues I_d and I_{ch} were selected from the range between -2.6 A and -7.8 A (for discharge) and between 1.3 A and 7.8 A (during the CC charge phase). The average charge current from all half-cycles completed from BoL ($SoH = 100\%$) to EoL ($SoH = 80\%$) was adopted as the value of charge current during analyses. During the experimental procedures, LICs were kept in the voltage operating window specified by the manufacturer, i.e. from 2.75 V (U_{dsch}) to 4.2 V (U_{ch}). Several DoD levels were selected ranging between 16% and 100%, with an assumption that DoD is the charge which can be obtained in the given combination of cycling parameters. For example $DoD = 100\%$ corresponds to a discharge in the full voltage operating window. The interval between charge and discharge half-cycles was constant and was equal to 15 s. The cell cycling process was controlled by a Cadex C8000 tester dedicated to electrochemical cells (values of momentary currents, voltages and temperatures of LICs were measured using separate wires). The current/voltage measurement accuracy amounted to ± 0.001 A/V while the temperature measurement accuracy was $\pm 0.1^\circ\text{C}$. To sum up, 29 cells in total were tested at four different T_a (10°C, 15°C, 25°C and 40°C), at 3 different values of I_d (-2.6 A, -5.2 A and -7.8 A) and at 5 different DoD levels (16%, 27%, 50%, 75% and 100%). The average value of charge current ($I_{ch, avg}$) ranged between 1.2 A and 3.46 A. Table I contains detailed values of the cycling parameters adopted during the aging tests.

The condition for completion of each aging procedure was the degradation of the cell to $SoH = 80\%$. The actual SoH of the cells was checked every 48 hours under reference conditions (described in algorithm 1) using the following relationship (1):

$$SoH = \frac{Q_{ref}}{Q_{BoL}} \cdot 100\% \quad (1)$$

TABLE I
VALUES OF CYCLING PARAMETERS SELECTED FOR AGING PROCEDURES AND RESULTS OBTAINED FROM INITIAL CALCULATIONS

Cell ID	T_c (°C)	I_d (A)	$I_{s, avg}$ (A)	DoD (%)	FEC _{int} (-)	RUE _{at BoL} (-)	RUE _{at EoL} (-)	ΔRUE_c (%)	$E_{in, int}$ (Wh)
H5	18.9	-2.6	2.08	100	83	0.872	0.687	18.5	2302
A31	29	-7.8	2.62	100	184	0.827	0.418	40.9	3451
D11	17.8	-2.6	2.47	50	183	0.545	0.545	0	3435
B37	20.9	-2.6	2.35	100	72	0.872	0.709	16.3	1516
A30	17.8	-2.6	2.4	73	149	0.776	0.663	11.3	2753
H2	16.6	-2.6	1.8	27	275	0.304	0.304	0	5029
A36	25.1	-5.2	2.09	50	426	0.536	0.536	0	8002
D14	23.9	-2.6	2.15	100	130	0.88	0.74	14	2517
B40	30.3	-5.2	2.2	100	1054	0.875	0.55	32.5	19804
A35	32.5	-7.8	2.52	100	664	0.865	0.33	53.5	12449
A20	25.7	-5.2	1.81	100	209	0.942	0.704	23.8	3933
C33	32.7	-2.6	3.04	74	306	0.789	0.749	4	5765
H10	33	-2.6	2.78	74	344	0.795	0.759	3.6	6478
H9	35.3	-2.6	2.85	100	226	0.945	0.752	19.3	4254
D16	43.2	-7.8	2.64	100	905	0.911	0.294	61.7	16978
B32	39.7	-5.2	2.66	100	676	0.922	0.672	25	13573
D1	33.2	-5.2	2.38	16	3984	0.165	0.165	0	74901
A24	36.6	-5.2	2.18	50	1721	0.546	0.546	0	32340
D4	29.2	-2.6	1.2	100	1908	0.982	0.786	19.6	36109
C26	36.4	-5.2	1.9	100	909	0.97	0.689	28.1	17083
A23	36.6	-7.8	1.74	100	557	0.939	0.221	71.8	10458
H12	58.6	-7.8	2.68	100	587	0.931	0.533	39.8	10993
H1	53.8	-7.8	1.88	100	563	1.005	0.504	50.1	10555
D6	50.1	-2.6	3.46	100	375	1.046	0.813	23.3	7048
B35	47.5	-2.6	2.37	100	702	1.057	0.848	20.9	13206
A38	52.5	-5.2	2.54	100	960	1.002	0.717	28.5	18034
H13	62.1	-7.8	3.11	75	716	0.744	0.564	18	13409
C10	32.6	-2.6	2.61	100	1460	0.995	0.798	19.7	27446
D2	33.7	-2.6	3.2	100	160	0.934	0.742	19.2	3028

where Q_{ref} is the collected charge, Q_{BoL} is the charge collected before the start of the LIC cycling.

2.2. Useful energy determination

In order to determine the useful energy throughput of a LIC during a single duty cycle over its entire life, the following methodology was adopted:

- 1) The assumption that one complete duty cycle consists of a discharge half-cycle and the immediately following charge half-cycle.
- 2) Calculation of the charge throughput during a single duty cycle (collected during discharge Q_{dsch} and charge Q_{ch}) using the Coulomb Counting method, on the basis of equation (2).

- 3) Determination of the relative useful energy RUE_c throughput during the duty cycle in accordance with relationship (3). $RUE_c=1$ means that a cell transferred nominal energy (18.876 Wh) during a full duty cycle (a discharge followed by a charge).
- 4) After each control procedure, determination of the number of completed full equivalent duty cycles (FECs) for the needs of the prediction model.
- 5) Determination of FEC_{tot} as the sum of RUE_c from all completed LIC duty cycles from BoL to EoL . An assumption was made that the transfer of nominal energy (18.876 Wh) during two consecutive charge and discharge half-cycles constitutes a full equivalent cycle (FEC).
- 6) Determination of the total energy throughput $E_{th,tot}$ as a product of total charge throughput of LIC, multiplied by its voltage on terminals:

$$Q_{ch/dsch} = Q_0 + \int_0^t I_{ch/dsch}(t) dt \quad (2)$$

$$RUE_c = \frac{Q_{dsch} \cdot U_{ave_dsch} + Q_{ch} \cdot U_{ave_ch}}{2 \cdot C_{nom} \cdot U_{nom}} \quad (3)$$

where Q_0 is the initial charge accumulated in a LIC, U_{ave_dsch} is an average voltage during a discharge half-cycle, U_{ave_ch} is an average voltage during a charge half-cycle, C_{nom} is the nominal capacity and U_{nom} is the nominal voltage.

2.3. Initial analysis and calculations

The tests carried out showed that for the LICs discharged to $DoD > 70\%$ the value of the transferred RUE_c is variable during their considered life. In the case of variants with $DoD < 70\%$, LICs transferred the fixed energy throughout their life ($RUE_c = const$). Depending on the combination of the values of the cycling parameters, three types of evolution of RUE_c changes were observed during the analysed period of operation: a) approximately linear, b) a slow decrease in the first phase of life followed by a phase of accelerated loss of energy transfer, c) a characteristic inflection point occurred after the phase of releasing the assumed charge, followed by a phase of rapid loss of energy transfer. The phenomena described in points b) and c) were present in cells "A23", "D14", "D16", "A30", "H10", "A20", "B37", "H5" and "C33". Moreover, it was observed that the lower the average temperature during full duty cycle (T_c), the more rapidly the process of RUE_c loss proceeded. For LICs with $T_c > 30^\circ C$ in most cases (except "B35", "D16" and "A23") this phenomenon proceed approximately linearly. The selected characteristics of the RUE_c of cells, in which the effect described above occurred, are presented in Fig. 2.

For each completed aging variant, it was checked how the ability of the cell to transfer RUE_c decreased. For this purpose, the ΔRUE_c difference (eq. 4) between RUE_c determined at BoL and EoL was calculated.

$$\Delta RUE_c = (RUE_{c_BoL} - RUE_{c_EoL}) \cdot 100\% \quad (4)$$

Depending on the cycling conditions ΔRUE_c reached values exceeding 50% (cells: "A35", "D16", "H1" and "H23") which means that in the period preceding EoL the cells were not capable of transferring even half of the initial RUE_c . A high value of ΔRUE_c was observed for cells discharged with a high current (-7.8 A) regardless of the temperature conditions (average cyclic temperature of the cell). The lowest values of ΔRUE_c occurred for cycled cells at incomplete DoD ("A30", "C33", "H10" and "H13").

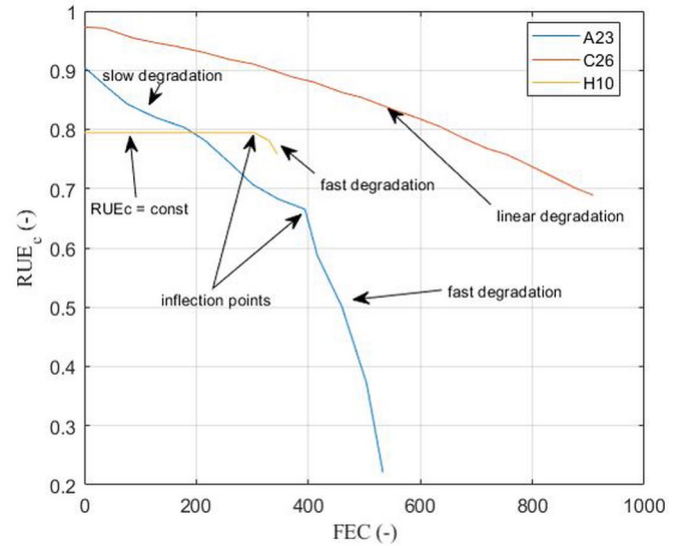


Fig. 2. Three types of evolution of RUE_c trends observed for the tested LIC

Additionally, FEC_{tot} and $E_{th,tot}$ completed by each LIC were determined and listed for information purposes in Table I. The statistical values of measuring data and initial calculations are included in Table II. Raw results of experimental measurements and FEC calculations of all LICs are available online (raw data: <https://data.mendeley.com/datasets/fzp5wx28kw/1>).

TABLE II
STATISTICAL VALUES OF THE DATA OBTAINED FROM EXPERIMENTAL AND INITIAL CALCULATIONS

Parameter	T_c ($^\circ C$)	I_d (A)	$I_{ch,avg}$ (A)	DoD (%)	RUE_c (-)
Missing values	0	0	0	0	0
Minimum	17.7	-2.6	1.213	73	0.221
Maximum	58.6	-7.8	3.46	100	1.057
Median	33	-2.6	2.4	100	0.823
Mean	36.9	-4.67	2.23	96.7	0.811
Std. dev.	10.41	2.41	0.603	8.69	0.14
Variance	108.3	5.82	0.364	75.4	0.02
25 % percentile	29.2	-7.8	1.88	100	0.757
75 % percentile	45.5	-2.6	2.6	100	0.907
Number of different values	19	3	18	3	211

3. Proposed methodology

3.1. Data pre-processing

An inseparable part of prediction models based on machine learning is data pre-processing. The techniques used in data pre-processing enable assessment of the quality of the data and their usefulness, which translates directly into the ability of the created model to learn. Due to the numerical nature of the data used, it was not necessary to use methods related to the handling of categorical variables and null values. Before initiating the machine learning procedure, the data obtained from the experimental measurements were normalised. After the normalisation, the average value of each of the parameters of the learning dataset was equal to 0, and the standard deviation was equal to 1.

3.2. Proposed model

The amount of RUE_c which LICs are capable of transferring during a full duty cycle over their lifetime depends on the combination of values of cyclic operating parameters and the number of completed $FECs$. As the tests demonstrated, LICs may have a similar ability to transfer RUE_c even if they have completed a significantly different number of $FECs$. One additional issue is the fact that the cyclic operating parameters of a LIC are strongly correlated. For instance higher values of DoD , I_d and $I_{ch,avg}$ result in an increase in the average temperature of the LIC, which, in turn, determines the amount of RUE_c . Furthermore, the impossibility of separating the impact of the above-mentioned parameters during a duty cycle makes it highly difficult to determine the precise effect of individual parameters on the amount of RUE_c . In this paper, in order to develop a prediction model for the relative useful energy RUE_c the framework presented in Fig. 3 was adopted.

Taking the above-mentioned challenges into account, the author proposes the use of the non-parametric model, whose structure strictly matches the form of learning data. This model belongs to the group of machine learning models (with supervision); it is more frequently referred to as GPR and is a non-parametric regression technique based on Gauss processes. This approach has many advantages, the most important being the lack of requirements to know the distributions of the model parameters and the correlations between them, and also a high prediction accuracy even when using a small learning dataset. Additionally, it forces out the use of techniques which enable the interpretation of the manner in which the implemented non-parametric model operations. These techniques belong to the so called explainable artificial intelligence XAI. In this paper, the author have applied a method that makes it possible to determine the significance of explanatory variables in the model. For the purpose of quantitative determination of the effect which the respective model variables have on the predicted value (taking into account the correlation of other parameters), the concept of accumulated local effects (ALE) has been used.

3.3. Gaussian process regression

Gaussian process regression (GPR) belongs to the group of kernel-based non-parametric models [42]. In this model, the Gaussian probability distribution is defined for each finite set of input variables x_i :

$$f(x) \sim GP(\mu(x), cov(x, x')) \quad (5)$$

where $\mu(x)$ is mean function, and $cov(x, x')$ is the covariance function (kernel function).

The mean and covariance functions are expressed as:

$$\mu(x) = E(f(x)) \quad (6)$$

$$cov(x, x') = E((f(x) - \mu(x))(f(x') - \mu(x')))) = k(x, x') \quad (7)$$

where E is an expected value.

The mean function may be equal to zero and may be a constant or mean value of the learning dataset. On the other hand, the kernel function may have various specific forms ranging from the standard ones, such as constant, linear, radial, squared exponential or matern, as well as being a composite of multiple kernel functions. The kernel functions contain parameters related to the scaling of responses of model y and input vector x and are referred to as hyperparameters θ . For the majority of standard kernels, the hyperparameters include: standard deviation σ_f and characteristic length σ_l . For all the parameters of the input vector of the model, the characteristic length may be identical or different.

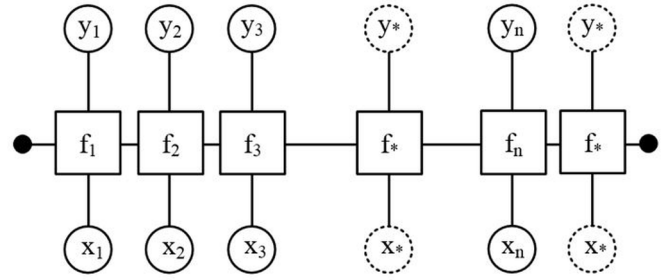


Fig. 4. Graphic representation of the GPR model

The matching of the structure of the model to the learning set takes place by defining, for each sample of the learning dataset x_i , a function $f(x_i)$ having the Gaussian distribution – hereafter called a hidden variable – and a set of basis functions $h(x)$, which transfer the input vector x_i from the original feature space R^d to the extended feature space R^p . The graphic representation of the GPR model is presented in Fig. 4. The known values (learning dataset) are marked in circles with continuous lines, while the hidden variables of the GPR model are mapped using rectangles. The hidden variables are connected by a thin horizontal line and, taking into account the marginalisation property of the model, each data sample (x_i, y_i) is independent of the others. Points (x^*, y^*) in circles with dashed lines mean new data points.

By using the kernel functions, it is possible to determine how model prediction y for vector x is dependent on the response in other points x' . It determines the similarity between the two points (x, x') in the scalar form $k(x, x')$. More details related to kernel functions can be found in [18].

For the needs of the RUE_c prediction, the following structure of the model has been adopted:

$$y(x) = h(x)^T w + f(x) + \varepsilon \quad (8)$$

where $h(x)$ is the set of basis functions, w is the vector of basis function coefficients, ε is noise with normal distribution.

Model prediction y takes place on the basis of input vector x' and the learning dataset:

$$P(y_i | f(x_i), x_i) \sim N(y_i | h(x_i)^T w + f(x_i), \sigma^2) \quad (9)$$

The above-mentioned model may be expressed in the form of a vector using the following relationship:

$$P(\mathbf{y} | \mathbf{f}, \mathbf{X}) \sim N(\mathbf{y} | \mathbf{H}\mathbf{w} + \mathbf{f}, \mathbf{I}\sigma^2) \quad (10)$$

where σ is the standard deviation, \mathbf{I}_n is the identity matrix and:

$$\mathbf{X} = (x_1^T \ x_2^T \ \dots \ x_n^T)^T \quad (11)$$

$$\mathbf{f} = (f(x_1) \ f(x_2) \ \dots \ f(x_n))^T \quad (12)$$

$$\mathbf{H} = (h(x_1^T) \ h(x_2^T) \ \dots \ h(x_n^T))^T \quad (13)$$

$$\mathbf{y} = (y_1 \ y_2 \ \dots \ y_n)^T \quad (14)$$

Joint probability distribution of latent variables \mathbf{f} with an assumption of a zero mean function $\mu(x)=0$ is expressed as follows:

$$P(\mathbf{f}|\mathbf{X}) \sim N(\mathbf{f}|0, \mathbf{K}(\mathbf{X}, \mathbf{X}')) \quad (15)$$

where:

$$\mathbf{K}(\mathbf{X}, \mathbf{X}') = \begin{pmatrix} k(x_1, x_1') & k(x_1, x_2') & \cdots & k(x_1, x_n') \\ k(x_2, x_1') & k(x_2, x_2') & \cdots & k(x_2, x_n') \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1') & k(x_n, x_2') & \cdots & k(x_n, x_n') \end{pmatrix} \quad (16)$$

The values of the hyperparameters of the kernel function can be determined by maximisation of the logarithmic membership function $P(\mathbf{y}|\mathbf{X})$ with respect to parameters \mathbf{w}, θ and σ^2 using the following relationship:

$$\mathbf{w}, \theta, \sigma^2 = \arg \max_{\mathbf{w}, \theta, \sigma^2} \log P(\mathbf{y}|\mathbf{X}, \mathbf{w}, \theta, \sigma^2) \quad (17)$$

Taking into account that:

$$P(\mathbf{y}|\mathbf{X}, \mathbf{w}, \theta, \sigma^2) = N(\mathbf{y}|\mathbf{H}\mathbf{w}, \mathbf{K}(\mathbf{X}, \mathbf{X}') + \mathbf{I}\sigma^2) \quad (18)$$

the logarithmic membership function may be represented as:

$$\log P(\mathbf{y}|\mathbf{X}, \mathbf{w}, \theta, \sigma^2) = -\frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{w})^T [\mathbf{K}(\mathbf{X}, \mathbf{X}') + \mathbf{I}\sigma^2]^{-1} (\mathbf{y} - \mathbf{H}\mathbf{w}) - \frac{n}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{K}(\mathbf{X}, \mathbf{X}') + \mathbf{I}\sigma^2| \quad (19)$$

After determining the parameters of the model, the probabilistic prediction of new values y_{new} for a new input vector x_{new} can take place as follows:

$$P(y_{new}|\mathbf{y}, \mathbf{X}, \mathbf{x}) = \frac{P(y_{new}, \mathbf{y}|\mathbf{X}, \mathbf{x}_{new})}{P(\mathbf{y}|\mathbf{X}, \mathbf{x}_{new})} \quad (20)$$

where:

$$P(y_{new}|\mathbf{y}, \mathbf{X}, \mathbf{x}_{new}) = \iint P(y_{new}|f_{new}, x_{new}) P(\mathbf{f}|\mathbf{y}, \mathbf{X}) P(f_{new}|\mathbf{f}, \mathbf{X}, \mathbf{x}_{new}) d\mathbf{f} df_{new} \quad (21)$$

The individual components of the above integral can be represented as follows:

$$P(y_{new}|f_{new}, x_{new}) = N(y_{new}|\mathbf{h}(\mathbf{x}_{new})^T \mathbf{w} + f_{new}, \sigma_{new}^2) \quad (22)$$

$$P(\mathbf{f}|\mathbf{y}, \mathbf{X}) = N\left(\mathbf{f} \left| \frac{1}{\sigma^2} \left(\frac{\mathbf{I}}{\sigma^2} + \mathbf{K}(\mathbf{X}, \mathbf{X}') \right)^{-1} (\mathbf{y} - \mathbf{H}\mathbf{w}), \left(\frac{\mathbf{I}}{\sigma^2} + \mathbf{K}(\mathbf{X}, \mathbf{X}') \right)^{-1} \right.\right) \quad (23)$$

$$P(f_{new}|\mathbf{f}, \mathbf{X}, \mathbf{x}_{new}) = N\left(f_{new} | \mathbf{K}(\mathbf{x}_{new}^T, \mathbf{X}) \mathbf{K}(\mathbf{X}, \mathbf{X}')^{-1} \mathbf{f}, \Delta\right) \quad (24)$$

where:

$$\Delta = k(x_{new}, x_{new}') - \mathbf{K}(\mathbf{x}_{new}^T, \mathbf{X}) \mathbf{K}(\mathbf{X}, \mathbf{X}')^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x}_{new}') \quad (25)$$

Finally the probability density of prediction y_{new} at the new point x_{new} with known \mathbf{y} and \mathbf{X} is given by the following equation:

$$P(y_{new}|\mathbf{y}, \mathbf{X}, \mathbf{x}_{new}) = N(y_{new}|\mathbf{h}(\mathbf{x}_{new})^T \mathbf{w} + \rho, \sigma_{new}^2 + \Sigma) \quad (26)$$

where:

$$\hat{\mathbf{n}} = \mathbf{K}(\mathbf{x}_{new}^T, \mathbf{X}) (\mathbf{K}(\mathbf{X}, \mathbf{X}') + \mathbf{I}\sigma^2)^{-1} (\mathbf{y} - \mathbf{H}\mathbf{w}) \quad (27)$$

$$\Sigma = k(x_{new}, x_{new}') - \mathbf{K}(\mathbf{x}_{new}^T, \mathbf{X}) (\mathbf{K}(\mathbf{X}, \mathbf{X}') + \mathbf{I}\sigma^2)^{-1} \mathbf{K}(\mathbf{X}, \mathbf{x}_{new}') \quad (28)$$

3.4. Variable importance

Variable importance is the technique which allows for the determination of the significance of each explanatory variable in reference to its effect on the implemented machine learning model [34,62]. This concept consists of the determination of the model error by permutations of the values of each variable. The given variable is significant if changes in its value cause changes in the model error, as the model relied on this variable for prediction. If the permutations of the variable do not cause changes in the model error, then it is considered insignificant. In other words, the more the model relies on a variable to make predictions, the more important it is for the model. The values of significance of the variables are determined on the basis of algorithm 2.

3.5. Accumulated local effects

Accumulated local effects (ALE) is the method which allows for the assessment of the relationship between the explanatory variables of the model and the predicted value [1,34]. This concept is based on conditional variable distributions. This means that mean differences in prediction and not their mean values are taken into account (as is the case with the partial dependence (PD) concept [34]). ALE reflects the way in which model forecasts change within the narrow range of values of variables. This method is particularly useful in the case of sets whose parameters are correlated with each other. The ALE function is defined as follows:

$$ALE(x_1) = \int_{\min(x_1)}^{x_1} E \left[\frac{\partial f(x_1, x_2)}{\partial x_1} \Big|_{x_1 = z_1} \right] dz_1 + c_1 = \int_{\min(x_1)}^{x_1} \int p(x_2|z_1) \frac{\partial f(z_1, x_2)}{\partial z_1} dx_2 dz_1 + c_1 \quad (29)$$

where partial derivative $\frac{\partial f(x_1, x_2)}{\partial x_1}$ represents the local effect of x_1 on prediction function f at (x_1, x_2) , $\min(x_1)$ is lower value of considered variable range, c_1 is a constant selected in order to vertically center the graph.

Averaging the local effect over the conditional distribution $p(x_2|x_1)$ makes it possible to separate the effect of other correlated variables located outside the considered range of values of variable x_1 .

In order to determine local effects, the whole range of values of the given explanatory variable must be divided into multiple intervals. The "effect" is understood as the difference in prediction calculated separately for each occurrence of a variable in a given interval. All differences in the given interval are summed up and divided by the number of occurrences – this way, the obtained mean difference in predictions is the "local" effect. Ultimately the value of the accumulated local effect of a given variable is determined by summing up all the local effects and represents the average change in the prediction when the value of the variable changes within its range.

Algorithm 2: Variable importance determination

Input:

X – variables matrix (explanatory variables)
 y – vector of predicted values of original data
 f – trained model
 $L(y, f)$ – loss function (any model performance method)

Steps:

1. Compute the original model error $L^0 = L(y, f(X))$
2. For each explanatory variable X_j :
 - a. In order to break the relationship between variable X_j and true prediction y – generate feature matrix X^p by permuting variable X_j in the data X .
 - b. Compute model predictions Y based on the permuted data
 - c. Calculate error based on the predictions of the permuted data $L_j = L(Y, f(X^p))$.
 - d. Calculate variable importance $VI_j = L_j - L^0$
3. Sort explanatory variables by descending VI

Output:

VI_j – variable importance of each explanatory variable

Numerically, ALE for a single variable can be calculated according to the following relationship:

$$ALE(x_1) = \sum_{k=1}^{k(x_1)} \frac{1}{n(k)} \sum_{i: x_1^{(i)} \in N(k)} [f(z_k, x_1^{(i)}) - f(z_{k-1}, x_1^{(i)})] - c_1 \quad (30)$$

where z_k represents the boundary of the range of variable x_1 , $N(k)$ is the interval $[z_{k-1}, z_k)$, $k(x_1)$ is the index of the interval to which a given point of variable x_1 belongs, $n(k)$, means the number of points within the interval under consideration, and f is the prediction function.

The ALE method also allows for analysis of changes in the predicted value in the case of interaction of two (ALE second-order) or more variables (ALE higher-order). In this case, the main impact of the variables concerned is ignored – only the effect of their interaction is calculated. For two explanatory variables, the ALE function may be represented as:

$$ALE(x_1, x_2) = \int_{\min(x_1)}^{x_1} \int_{\min(x_2)}^{x_2} E \left[\frac{\partial f(x_1, x_2, x_{\{1,2\}})}{\partial x_1 \partial x_2} \Big|_{x_1 = z_1, x_2 = z_2} \right] dz_1 dz_2 \quad (31)$$

and in the numerical form, the above relationship may be presented as follows:

$$ALE(x_1, x_2) = \sum_{k=1}^{k(x_1)} \sum_{m=1}^{k(x_2)} \frac{1}{n_{\{1,2\}}(k, m)} \sum_{i: x_{\{1,2\}}^{(i)} \in N(k, m)} \Delta f_{1,2}(K, k, m, x_{\{1,2\}}^{(i)}) \quad (32)$$

where:

$$\Delta f_{1,2}(K, k, m, x_{\{1,2\}}^{(i)}) = \left[f(z_{k,1}, z_{m,2}, x_{\{1,2\}}^{(i)}) - f(z_{k-1,1}, z_{m,2}, x_{\{1,2\}}^{(i)}) \right] - \left[f(z_{k,1}, z_{m-1,2}, x_{\{1,2\}}^{(i)}) - f(z_{k-1,1}, z_{m-1,2}, x_{\{1,2\}}^{(i)}) \right] \quad (33)$$

The rules for the determination of ALE for two variables are identical to those for one variable, whereby, the intervals are replaced by rectangular areas due to the necessity of accumulation of local effects

in two dimensions. ALE can also be calculated for the interaction of three or more variables – the relevant relationship can be found in paper [1].

4. Results and discussion

The input parameters of the model were the cyclic operating parameters: the average LIC temperature during the full duty cycle (T_c), the value of current during the discharge half-cycle (I_d), depth of discharge (DoD), the value of mean current during the charge half-cycle ($I_{ch, avg}$). The total aging processes leading to LIC degradation were included in the model by the number of full equivalent cycles (FEC). The predicted value was the relative useful energy (RUE_c) throughput of a cell during a full duty cycle.

4.1. Model training

The results of measurements of cycling of LICs, which were cycled with $DoD > 70\%$ (20 LICs in total), were used as the learning dataset. Cells "A38", "B32", "B40" and "H9" were used to verify the model exclusively. Finally, the learning dataset contained 311 samples, and the verification set consisted of 99 samples – a fragment of the learning dataset after initial calculations is presented in Table III. The value of $FEC=0$ is equivalent to $SOH=100\%$ of the cell, while the last value of $FEC=533$ indicates $SOH=80\%$ and the accomplishment of the EoL status.

As the kernel function of the GPR model, the matern 3/2 function was used, which is expressed by the following equation:

$$k_{mat3/2}(x, x') = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{\sigma_l} \right) \exp \left(-\frac{\sqrt{3}r}{\sigma_l} \right) \quad (34)$$

where r is Euclidean distance between x and x' .

TABLE III
SAMPLE OF LEARNING DATASET – CELL „A23”

T_c (°C)	I_d (A)	DoD (%)	$I_{ch, avg}$ (A)	FEC (-)	RUE_c (-)
36.6	-7.8	100	1.74	0	0.905
36.6	-7.8	100	1.74	41	0.871
⋮	⋮	⋮	⋮	⋮	⋮
36.6	-7.8	100	1.74	504	0.371
36.6	-7.8	100	1.74	533	0.221

The values of characteristic parameters σ_f and σ_l were determined using the method of optimisation without constraints (quasi-Newton) and were 0.1687 and 0.7659 respectively, while the value of the logarithmic membership function was $L=713.7$. The linear function type was adopted for the function set $h(x)$. The coefficient of determination R^2 , mean average percentage error $MAPE$ and root mean square error $RMSE$ were used to assess the model quality. The above indicators are determined according to the following relationships:

$$R^2 = \frac{\sum_{i=1}^n (f_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (35)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - f_i}{y_i} \right| \cdot 100\% \quad (36)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2} \quad (37)$$

where n is the number of samples, f is the predicted value, y is the measured value.

The results of the learning process were compared with three supervised machine learning models: a regression tree, a support vector machine (*SVM*), and an ensemble of regression trees and are summarized in Table IV. The same input parameter structure was used for each model, and the predictor parameter was RUE_c . For the regression tree, the lowest $RMSE$ was achieved for a leaf size of 4. It was observed that increasing the leaf size beyond this value resulted in an increase in $RMSE$. For the *SVM*-based model, the lowest $RMSE$ was obtained for a Gaussian-type kernel with a characteristic length of 0.56. The model with second lowest $RMSE$ was the ensemble of the regression trees consisting of a combination of 100 trees with a leaf size of 5 each and a learning rate of 0.25.

The example characteristics of the three observed types of RUE_c evolution for two models with lowest $RMSE$ (*GPR* model and Ensemble of trees model) are presented in Fig. 5 (a-c). The learning results of other LICs for all comparative models are included in Appendix A.

As opposed to the model based on Ensemble of regression trees, the response of the *GPR* model for each LIC included in the learning dataset was within the assumed confidence interval (95%). The *MAPE* of the *GPR* model developed for all samples of the learning dataset was 0.05%. This proves that the *GPR* model almost perfectly matches the learning dataset.

4.2. Prediction

The developed original model was verified using new load profiles which were not used to train the model. LICs with different mean

TABLE IV
LEARNING PROCEDURE – SUMMARY OF RESULTS FOR DIFFERENT MODELS

Model	MAPE (%)	R ² (-)	RMSE (-)
Ensemble of trees	0.73	0.95	0.0311
<i>GPR</i>	0.05	0.99	0.0141
Regression tree	3.54	0.85	0.0551
<i>SVM</i>	2.72	0.89	0.0465

cycling temperatures (T_c) at mean load current of $I_d = -5.2$ A were selected for verification. For instance, for $T_c > 40^\circ\text{C}$, the learning dataset included LICs loaded with $I_d = -2.6$ A and $I_d = -7.8$ A (cells "D6" and "H12"), therefore, the "A38" cell was selected for verification at higher temperatures. The verification process consisted of the RUE_c prediction by a model of 4 cells in the period between *BoL* and *EoL*, and of comparison of their values obtained during the experiment. The verification results are presented in Table V, and the characteristics for the 4 tested cases are presented in Fig. 6. From the results obtained it can be observed that in two cases ("H9" and "B40") *APE* exceeded 5%. The reason for obtaining higher *APE* for the indicated LICs is primarily due to the low number of samples in the learning dataset containing conditions similar to those under which they were cycled. For the "H9" LIC the problem is in the near *EoL* phase. During the charging half-cycle, the average charging current for this LIC was 2.85 A. The learning dataset contains 29 samples (9.3% of all samples) with similar values of cycling parameters (especially with high average charging currents) to "H9" cell. Cell "B40" performed a high number of *FECs* (over 1000) while in the learning dataset only 7% of all samples are with similar values (mainly coming from "C26" LIC, which performed 909 *FEC*). This compares to 14.1% of samples

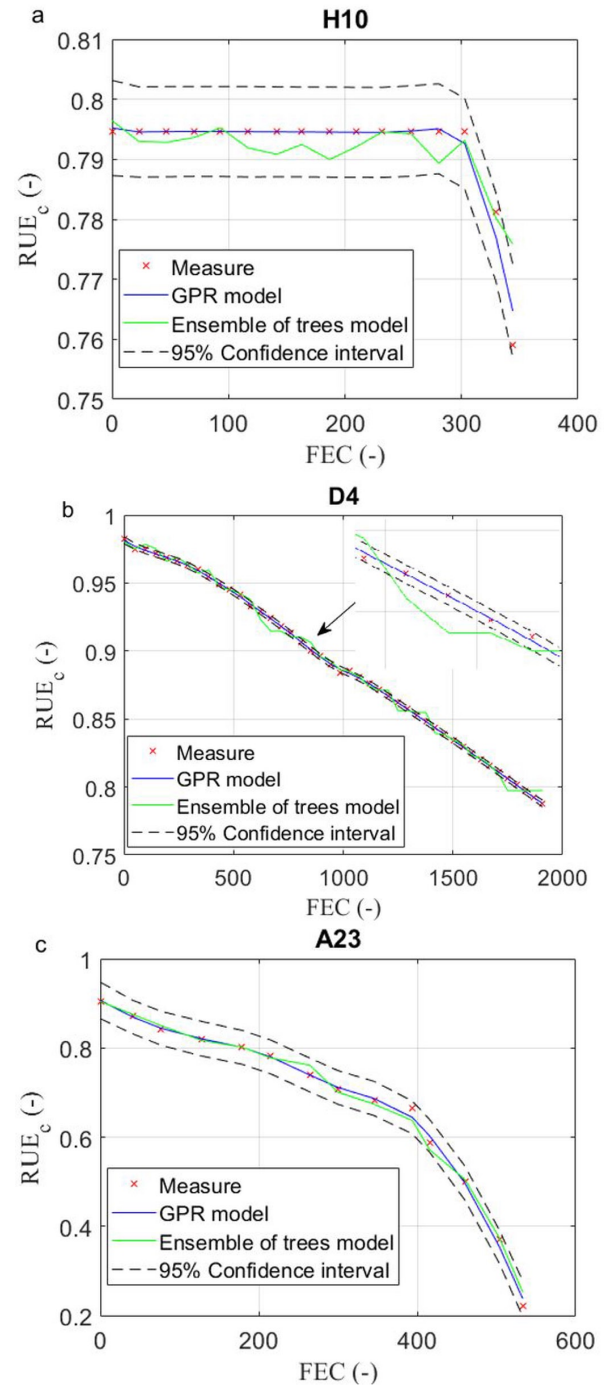


Fig. 5. Learning results obtained for three observed types of RUE_c evolution of *GPR* model and Ensemble of regression trees: a) after about 300 *FECs*, the cell lost the ability to transfer the fixed energy – the phase of significant decrease in RUE_c takes place, b) under these conditions, throughout the lifetime of the LIC, the decrease in RUE_c was approximately linear, c) the phase of slow degradation occurs until approximately 400 *FECs* are completed, then it is possible to observe the phase of accelerated ability to transfer RUE_c

with similar values in the learning dataset for cell "A38" and 14.5% for "B32".

In the author's opinion, the obtained results of the verification process confirm the effectiveness of the selected *GPR* method for the purpose of RUE_c prediction which a LIC can transfer during a full duty cycle throughout its life. Only in one test case did the *MAPE* slightly exceed 5% (the "B40" cell). Also in favor of the method used, is the fact that despite an almost ideal matching of the model to the learning data, the prediction of RUE_c under new operating conditions for the LIC is possible with little error.

The GPR model also allows for prediction within ranges exceeding the ranges of variables of the learning dataset, which is not possible for some techniques based on artificial intelligence – for instance, neural networks or fuzzy logic.

4.3. Model explanation

In order to determine the significance of the explanatory variables of the model, the variable importance (VI) technique was used. *RMSE* was selected as the loss function of *L* model (eq. 37). The calculations were repeated 8 times in order to negate the effect of uncertainty caused by permutations of values. The obtained results of VI are characterized by a low standard deviation; they were averaged and listed in Table VI. Based on the results, the respective model input parameters were ordered in terms of their significance - the effect of *FEC* on the predicted value of *RUE_c* is the highest in the used dataset (the highest VI), then the effect of *I_d* and *T_c* is on a comparable level. From among the adopted model parameters, *DoD* and *I_{ch_avg}* have the lowest effect on *RUE_c* (in both cases the calculated significance is at a similar level).

Given the fact that interpretation of the structure of the developed GPR non-parametric model is very difficult, in order to determine the effect which the respective model parameters have on the value of

TABLE V
RESULTS OF VERIFICATION PROCEDURE ON NEW LOAD PROFILES

Cell ID	<i>T_c</i> (°C)	<i>I_d</i> (A)	<i>DoD</i> (%)	<i>I_{ch_avg}</i> (A)	<i>MAPE</i> (%)
B40	30.3	-5.2	100	2.2	5.63
H9	35.3	-2.6	100	2.85	2.88
A38	54.6	-5.2	100	2.63	2.77
B32	39.7	-5.2	100	2.47	3.26

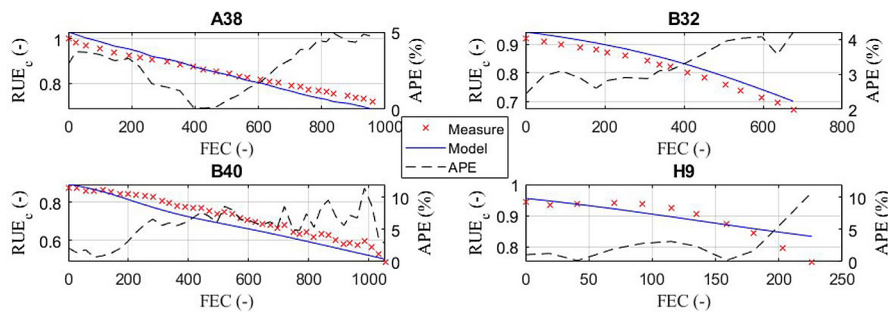


Fig. 6. Results of the model verification procedure

TABLE VI
VALUES OF VARIABLES IMPORTANCE OF IMPLEMENTED MODEL

Parameter	<i>T_c</i> (°C)	<i>I_d</i> (A)	<i>DoD</i> (%)	<i>I_{ch_avg}</i> (A)	<i>FEC</i> (-)
mean <i>VI</i>	0.120042	0.147175	0.048705	0.056989	0.167583
std. dev.	0.001375	0.001651	0.000866	0.000554	0.001614

the *RUE_c* prediction, the author has applied the ALE concept. Fig. 7 (a-e) presents the plots of the first-order accumulated local effects which determine the impact of individual input parameters of the GPR model on *RUE_c*. The value of ALE at the given point determines the difference in prediction in relation to the mean prediction of *RUE_c*. ALE=0 represents the mean prediction of *RUE_c*. For instance, for *T_c* ≈ 25°C, the *RUE_c* prediction will be lower by 0.05 in relation to the mean prediction which is an exclusive consequence of temperature (the effect of other cyclic operating parameters is separated). Based

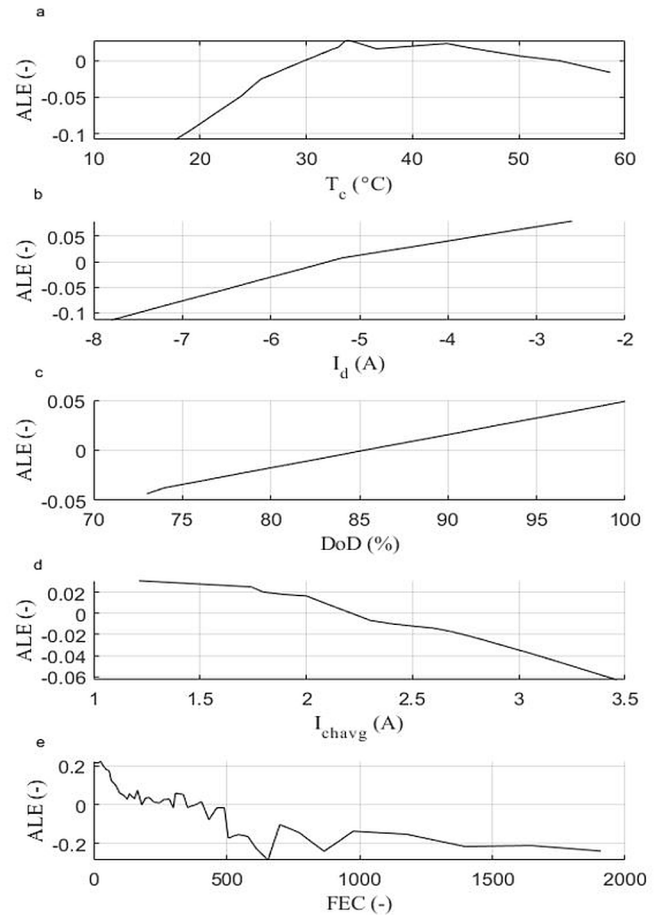


Fig. 7. First-order ALE plots for individual model input parameters: a) average temperature of LIC during the full cycle b) discharge current c) depth of discharge d) average charge current e) full equivalent cycle.

on the results obtained, it is possible to determine, unambiguously, the effect of *T_c* on *RUE_c* – the lowest values of *RUE_c* are obtained at average cycling temperatures below 30°C and they increase gradually up to the range of values between 30°C and 44°C, where *RUE_c* reaches its peak value and then falls for *T_c* higher than 44°C (Fig. 7 – a). In the case of values of currents *I_d* and *I_{ch_avg}*, the effect on *RUE_c* is linear. The lower the values of *I_d* and *I_{ch_avg}* are, the higher the *RUE_c* is. Furthermore, cycling of the LICs at *I_d* values lower than -5.2 A and higher than 2.25 A for *I_{ch_avg}* will result in lower *RUE_c* values over their lifetime – the visible breaking point of characteristics (Fig. 7 – b, d). The plot of ALE for *DoD* showed the linear dependence of this factor on *RUE_c* (Fig. 7 – c). The amount of transferred *RUE_c* drops sharply when the cells complete more than 500 full equivalent duty cycles (Fig. 7 – e).

Additionally, the results obtained using the ALE method confirm the results obtained using the VI technique – the highest dispersion of ALE values (from -0.286 to 0.225) is characteristic of *FEC*, which is compliant with the highest value of VI for this parameter. For *I_d* and *T_c* the differences in relation to the mean prediction vary from -0.114 to 0.08 for *I_d* and from -0.108 to 0.027 for *T_c* respectively, which also correlates with the VI results. The lowest dispersion of values was observed for *I_{ch_avg}* – from -0.063 to 0.031, which means the lowest effect on *RUE_c* in the used dataset.

In order to determine the effect of interaction of the model parameters on *RUE_c*, the values of the second-order accumu-

lated local effects were determined in the paper for all the possible combinations of pairs of the GPR model input parameters. The results obtained in the form of value maps are presented in Fig. 8 (a–j). Analysis of the interaction effect of model parameter pairs showed that the highest influence on RUE_c is exerted by combinations of the following pairs T_c – FEC , I_d – FEC and I_{ch_avg} – FEC . For instance, for LICs, which completed a large number of $FECs$ (over 1500) at T_c below 25°C, the differences in relation to the mean RUE_c prediction even amount to -0.4, and at T_c above 50°C +0.2 respectively (Fig. 8 a). A high level of influence was also shown by the I_{ch_avg} – FEC pair at around $FEC=1350$ and $I_{ch_avg}=3A$ – the ALE value in this area reaches up to +0.3 (Fig. 8 d). The level of influence of the DoD – FEC pair over the entire value range does not exceed the range between -0.04 and +0.06 (Fig. 8 d). To sum up – the pairs containing FEC are characterised by the highest influence due to the highest effect of FEC on RUE_c which is shown in Fig. 7 – e. The interaction effect of parameter pairs, i.e. T_c – DoD and T_c – I_d is locally at the level of hundredths (+0.02 and -0.015) – Fig. 8 – h,j. On the other hand, the effect of other pairs is at a negligible level (ALE at a level of thousandths or lower) – Fig. 8 – e-g,i.

5. Conclusion

The paper presents the author's prediction model for the useful energy that can be transferred during a single duty cycle under various temperature and load conditions, taking into account the energy throughput to date by $LiNi_{0.33}Mn_{0.33}Co_{0.33}O_2$ cells.

APPENDIX

A. Results of the learning procedure – $MAPE$ of comparative models for individual LICs.

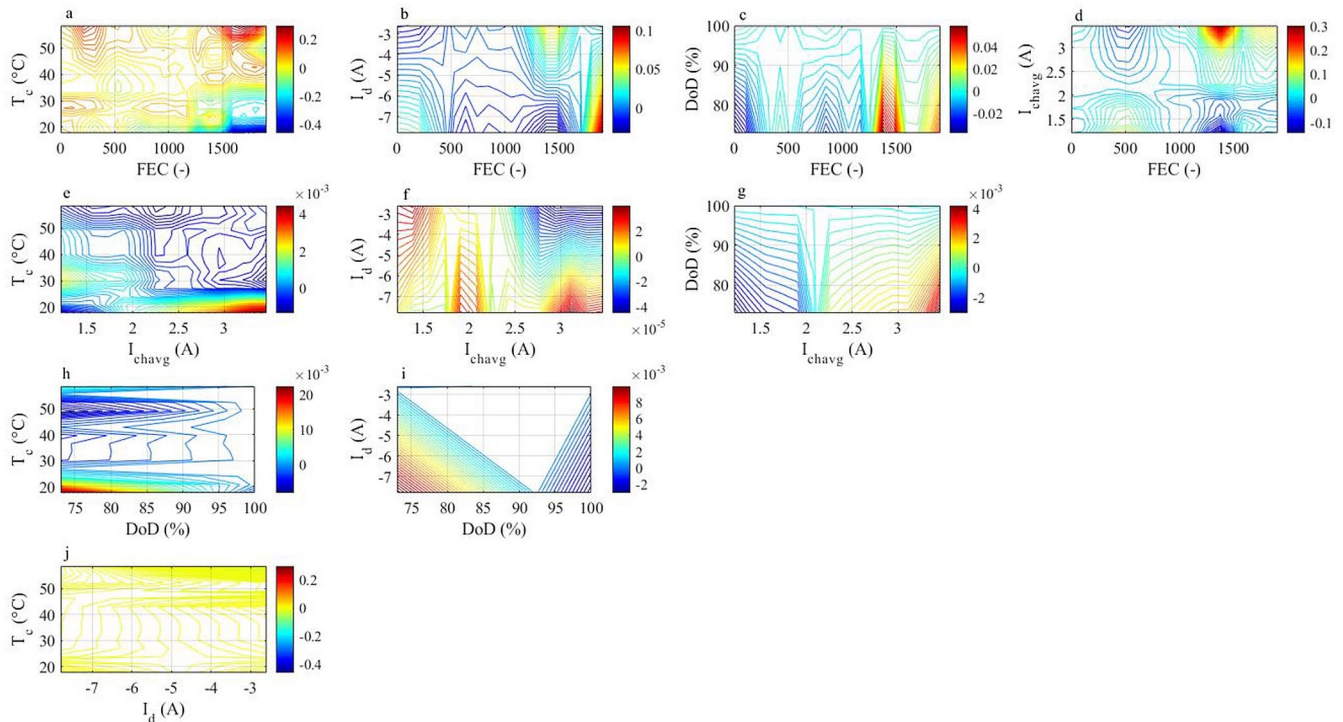


Fig. 8. Plots of accumulated local effects of the second order for the respective combinations of pairs of model input parameters

The methodology presented in the paper allows for the obtaining of a RUE_c prediction over the entire LIC life with high accuracy, even when learning datasets with a small number of samples are used. The developed model is resistant to the overfitting phenomenon – despite an almost ideal matching of the model to the learning dataset used, the prediction of trends in the evolution of RUE_c is possible under new load and temperature conditions, which has been confirmed by the author with 4 test cases. The appropriate selection of types for the kernel function $k(x, x')$ and basis functions $h(x)$ was critical to achieving the high accuracy of model prediction.

In comparison to classic modelling techniques, the advantage is that the model structure does not need to be specified. In fact, this is the disadvantage of methods based on machine learning – the structures of such models and their parameters have no physical meaning. The author proposes a solution for the above-mentioned problem by using the XAI techniques, which enable determination of both the significance of model input parameters, and the quantitative determination of their effect on the predicted value of RUE_c .

The information obtained in this way may be used, among other things, to improve the algorithms used specifically in battery or battery pack management systems, helping to optimise their operation and extend their life. What is more, the methodology proposed by the author can be applied to electrochemical cells with different chemical compositions.

References

1. Apley D W, Zhu J. Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 2020; 82(4): 1059-1086, <https://doi.org/10.1111/rssb.12377>.
2. Belt J, Utgikar V, Bloom I. Calendar and PHEV cycle life aging of high-energy, lithium-ion cells containing blended spinel and layered-oxide cathodes. *Journal of Power Sources* 2011; 196(23): 10213-10221, <https://doi.org/10.1016/j.jpowsour.2011.08.067>.
3. Bloom I, Cole B W, Sohn J J et al. An accelerated calendar and cycle life study of Li-ion cells. *Journal of Power Sources* 2001; 101(2): 238-247, [https://doi.org/10.1016/S0378-7753\(01\)00783-2](https://doi.org/10.1016/S0378-7753(01)00783-2).
4. Broussely M, Biensan P, Bonhomme F et al. Main aging mechanisms in Li ion batteries. *Journal of Power Sources* 2005; 146(1-2): 90-96, <https://doi.org/10.1016/j.jpowsour.2005.03.172>.
5. Burzyński D, Kasprzyk L. A novel method for the modeling of the state of health of lithium-ion cells using machine learning for practical applications. *Knowledge-Based Systems* 2021; 219: 106900, <https://doi.org/10.1016/j.knosys.2021.106900>.
6. Burzyński D, Pietracho R, Kasprzyk L, Tomczewski A. Analysis and Modeling of the Wear-Out Process of a Lithium-Nickel-Manganese-Cobalt Cell during Cycling Operation under Constant Load Conditions. *Energies* 2019; 12(20): 3899, <https://doi.org/10.3390/en12203899>.
7. Chen Z, Ren Y, Jansen A N et al. New class of nonaqueous electrolytes for long-life and safe lithium-ion batteries. *Nature Communications* 2013; 4: 1-8, <https://doi.org/10.1038/ncomms2518>.
8. Diao W, Saxena S, Pecht M. Accelerated cycle life testing and capacity degradation modeling of LiCoO₂-graphite cells. *Journal of Power Sources* 2019; 435(June): 226830, <https://doi.org/10.1016/j.jpowsour.2019.226830>.
9. Dong G, Zhang X, Zhang Ch, Chen Z. A method for state of energy estimation of lithium-ion batteries based on neural network model. *Energy* 2015; 90(1)1:879-888, <https://doi.org/10.1016/j.energy.2015.07.120>.
10. dos Reis G, Strange C, Yadav M, Li S. Lithium-ion battery data and where to find it. *Energy and AI* 2021, <https://doi.org/10.1016/j.egyai.2021.100081>.
11. Dudézert C, Reynier Y, Duffault J M, Franger S. Fatigue damage approach applied to Li-ion batteries ageing characterization. *Materials Science and Engineering B: Solid-State Materials for Advanced Technology* 2016; 213: 177-189, <https://doi.org/10.1016/j.mseb.2016.04.017>.
12. Ecker M, Gerschler J B, Vogel J et al. Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data. *Journal of Power Sources* 2012; 215: 248-257, <https://doi.org/10.1016/j.jpowsour.2012.05.012>.
13. Edström K, Dominko R, Fichtner M et al. BATTERY 2030+ Roadmap. 2020: 83, <https://doi.org/10.33063/diva2-1452023>.
14. El Mejdoubi A, Chaoui H, Gualous H et al. Lithium-ion batteries health prognosis considering aging conditions. *IEEE Transactions on Power Electronics* 2019; 34(7): 6834-6844, <https://doi.org/10.1109/TPEL.2018.2873247>.
15. Gao Y, Jiang J, Zhang C et al. Lithium-ion battery aging mechanisms and life model under different charging stresses. *Journal of Power Sources* 2017; 356: 103-114, <https://doi.org/10.1016/j.jpowsour.2017.04.084>.
16. Han X, Ouyang M, Lu L, Li J. A comparative study of commercial lithium ion battery cycle life in electric vehicle: Capacity loss estimation. *Journal of Power Sources* 2014; 268: 658-669, <https://doi.org/10.1016/j.jpowsour.2014.06.111>.
17. Hasib S A, Islam S, Chakraborty R K et al. A Comprehensive Review of Available Battery Datasets, RUL Prediction Approaches, and Advanced Battery Management. *IEEE Access* 2021; 9: 86166-86193, <https://doi.org/10.1109/ACCESS.2021.3089032>.
18. Haykin S, Simon S. *Neural Networks and learning machines*. Prentice Hall; New York 2009.
19. How D N T, Hannan M A, Lipu M S H et al. State-of-Charge Estimation of Li-Ion Battery in Electric Vehicles: A Deep Neural Network Approach. *IEEE Transactions on Industry Applications* 2020; 56(5): 5565-5574, <https://doi.org/10.1109/TIA.2020.3004294>.
20. Hu X, Jiang J, Cao D, Egardt B. Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling. *IEEE Transactions on Industrial Electronics* 2016; 63(4): 2645-2656.
21. Kasprzyk L. Modelling and analysis of dynamic states of the lead-acid batteries in electric vehicles. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2017; 19(2): 229-236, <https://doi.org/10.17531/ein.2017.2.10>.
22. Kasprzyk L, Domeracka A, Burzyński D. Modelowanie pracy i trwałości akumulatorów litowo-jonowych w pojazdach elektrycznych. *Przegląd Elektrotechniczny* 2018; 1(12): 160-163, <https://doi.org/10.15199/48.2018.12.34>.
23. Keil P, Schuster S F, Wilhelm J et al. Calendar Aging of Lithium-Ion Batteries. *Journal of The Electrochemical Society* 2016; 163(9): A1872-A1880, <https://doi.org/10.1149/2.0411609jes>.
24. Li I H, Wang W Y, Su S F, Lee Y S. A merged fuzzy neural network and its applications in battery state-of-charge estimation. *IEEE Transactions on Energy Conversion* 2007; 22(3): 697-708, <https://doi.org/10.1109/TEC.2007.895457>.
25. Li K, Wei F, Tseng K J, Soong B H. A Practical Lithium-Ion Battery Model for State of Energy and Voltage Responses Prediction Incorporating Temperature and Ageing Effects. *IEEE Transactions on Industrial Electronics* 2018; 65(8): 6696-6708, <https://doi.org/10.1109/TIE.2017.2779411>.
26. Li X, Miao J, Ye J. Lithium-ion battery remaining useful life prediction based on grey support vector machines. *Advances in Mechanical Engineering* 2015; 7(12): 1-8, <https://doi.org/10.1177/1687814015622327>.
27. Li X, Pan K, Fan G et al. A physics-based fractional order model and state of energy estimation for lithium ion batteries. Part II: Parameter identification and state of energy estimation for LiFePO₄ battery. *Journal of Power Sources* 2017; 367: 202-213, <https://doi.org/10.1016/j.jpowsour.2017.09.048>.
28. Li Y, Liu K, Foley A M et al. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renewable and Sustainable Energy Reviews* 2019. doi:10.1016/j.rser.2019.109254, <https://doi.org/10.1016/j.rser.2019.109254>.
29. Li Y, Li K, Liu X, Zhang L. Fast battery capacity estimation using convolutional neural networks. *Transactions of the Institute of Measurement and Control* 2020, <https://doi.org/10.1177/0142331220966425>.
30. Liu G, Ouyang M, Lu L et al. A highly accurate predictive-adaptive method for lithium-ion battery remaining discharge energy prediction in electric vehicle applications. *Applied Energy* 2015; 149: 297-314, <https://doi.org/10.1016/j.apenergy.2015.03.110>.
31. Liu K, Li Y, Hu X et al. Gaussian Process Regression with Automatic Relevance Determination Kernel for Calendar Aging Prediction of Lithium-Ion Batteries. *IEEE Transactions on Industrial Informatics* 2020; 16(6): 3767-3777, <https://doi.org/10.1109/TII.2019.2941747>.
32. Liu K, Shang Y, Ouyang Q, Widanage W D. A Data-Driven Approach with Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-ion Battery. *IEEE Transactions on Industrial Electronics* 2021; 68(4): 3170-3180, <https://doi.org/10.1109/TIE.2020.2973876>.

33. Liu T, Cheng L, Pan Z, Sun Q. Cycle life prediction of lithium-ion cells under complex temperature profiles. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2016; 18(1): 25-31, <https://doi.org/10.17531/ein.2016.1.4>.
34. Molnar C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. Book 2019: 247.
35. Motapon S N, Lachance E, Dessaint L-A, Al-Haddad K. A Generic Cycle Life Model for Lithium-Ion Batteries Based on Fatigue Theory and Equivalent Cycle Counting. *IEEE Open Journal of the Industrial Electronics Society* 2020; 1(August): 207-217, <https://doi.org/10.1109/OJIES.2020.3015396>.
36. Naumann M, Spingler F, Jossen A. Analysis and modeling of cycle aging of a commercial LiFePO₄/graphite cell. *Journal of Power Sources* 2020; 451(December 2019): 227666, <https://doi.org/10.1016/j.jpowsour.2019.227666>.
37. Ning G, Popov B N. Cycle Life Modeling of Lithium-Ion Batteries. *Journal of The Electrochemical Society* 2004; 151(10): A1584, <https://doi.org/10.1149/1.1787631>.
38. Olmos J, Gandiaga I, Saez-de-Ibarra A et al. Modelling the cycling degradation of Li-ion batteries: Chemistry influenced stress factors. *Journal of Energy Storage* 2021; 40(May): 102765, <https://doi.org/10.1016/j.est.2021.102765>.
39. Pan R, Wang Y, Zhang X et al. Power capability prediction for lithium-ion batteries based on multiple constraints analysis. *Electrochimica Acta* 2017; 238: 120-133, <https://doi.org/10.1016/j.electacta.2017.04.004>.
40. Pielecha I, Cieślak W, Szalek A. Operation of electric hybrid drive systems in varied driving conditions. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2018; 20(1): 16-23, <https://doi.org/10.17531/ein.2018.1.3>.
41. Pietracho R, Wenge C, Balischewski S et al. Potential of Using Medium Electric Vehicle Fleet in a Commercial Enterprise Transport in Germany on the Basis of Real-World GPS Data. *Energies* 2021; 14(17): 1-23, <https://doi.org/10.3390/en14175327>.
42. Rasmussen C E, Williams Ch K. Gaussian processes for machine learning. MIT press Cambridge; 2006:1, <https://doi.org/10.7551/mitpress/3206.001.0001>.
43. Richardson R R, Birkl C R, Osborne M A, Howey D A. Gaussian Process Regression for in Situ Capacity Estimation of Lithium-Ion Batteries. *IEEE Transactions on Industrial Informatics* 2019; 15(1): 127-138, <https://doi.org/10.1109/TII.2018.2794997>.
44. Richardson R R, Osborne M A, Howey D A. Battery health prediction under generalized conditions using a Gaussian process transition model. *Journal of Energy Storage* 2019; 23(March): 320-328, <https://doi.org/10.1016/j.est.2019.03.022>.
45. Safari M, Morcrette M, Teyssot A, Delacourt C. Multimodal Physics-Based Aging Model for Life Prediction of Li-Ion Batteries. *Journal of The Electrochemical Society* 2009; 156(3): A145, <https://doi.org/10.1149/1.3043429>.
46. Safari M, Morcrette M, Teyssot A, Delacourt C. Life Prediction Methods for Lithium-Ion Batteries Derived from a Fatigue Approach. *Journal of The Electrochemical Society* 2010; 157(7): A892, <https://doi.org/10.1149/1.3432560>.
47. Saha B, Goebel K, Poll S, Christophersen J. Prognostics methods for battery health monitoring using a Bayesian framework. *IEEE Transactions on Instrumentation and Measurement* 2009; 58(2): 291-296, <https://doi.org/10.1109/TIM.2008.2005965>.
48. Schmalstieg J, Käbitz S, Ecker M, Sauer D U. A holistic aging model for Li(NiMnCo)O₂ based 18650 lithium-ion batteries. *Journal of Power Sources* 2014; 257: 325-334, <https://doi.org/10.1016/j.jpowsour.2014.02.012>.
49. Schulze M C, Neale N R. Half-Cell Cumulative Efficiency Forecasts Full-Cell Capacity Retention in Lithium-Ion Batteries. *ACS Energy Letters* 2021; 6(3): 1082-1086, <https://doi.org/10.1021/acsenenergylett.1c00173>.
50. Sikha G, Popov B N, White R E. Effect of Porosity on the Capacity Fade of a Lithium-Ion Battery. *Journal of The Electrochemical Society* 2004; 151(7): A1104, <https://doi.org/10.1149/1.1759972>.
51. Su C, Chen H, Wen Z. Prediction of remaining useful life for lithium-ion battery with multiple health indicators. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2021; 23(1): 176-183, <https://doi.org/10.17531/ein.2021.1.18>.
52. Tagade P, Hariharan K S, Ramachandran S et al. Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis. *Journal of Power Sources* 2020; 445(October 2019): 227281, <https://doi.org/10.1016/j.jpowsour.2019.227281>.
53. Tao L, Ma J, Cheng Y et al. A review of stochastic battery models and health management. *Renewable and Sustainable Energy Reviews* 2017; 80(February 2018): 716-732, <https://doi.org/10.1016/j.rser.2017.05.127>.
54. Vetter J, Novák P, Wagner M R et al. Ageing mechanisms in lithium-ion batteries. *Journal of Power Sources* 2005; 147(1-2): 269-281, <https://doi.org/10.1016/j.jpowsour.2005.01.006>.
55. Vidal C, Malysz P, Kollmeyer P, Emadi A. Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art. *IEEE Access* 2020; 8: 52796-52814, <https://doi.org/10.1109/ACCESS.2020.2980961>.
56. Wang A, Kadam S, Li H et al. Review on modeling of the anode solid electrolyte interphase (SEI) for lithium-ion batteries. *npj Computational Materials* 2018, <https://doi.org/10.1038/s41524-018-0064-0>.
57. Wang J, Liu P, Hicks-Garner J et al. Cycle-life model for graphite-LiFePO₄ cells. *Journal of Power Sources* 2011; 196(8): 3942-3948, <https://doi.org/10.1016/j.jpowsour.2010.11.134>.
58. Wang J, Purewal J, Liu P et al. Degradation of lithium ion batteries employing graphite negatives and nickel e cobalt e manganese oxide b spinel manganese oxide positives : Part 1 , aging mechanisms and life estimation. *Journal of Power Sources* 2014; 269: 937-948, <https://doi.org/10.1016/j.jpowsour.2014.07.030>.
59. Wang Y, Zhang C, Chen Z. A method for joint estimation of state-of-charge and available energy of LiFePO₄ batteries. *Applied Energy* 2014; 135: 81-87, <https://doi.org/10.1016/j.apenergy.2014.08.081>.
60. Wang Y, Zhang C, Chen Z. Model-based state-of-energy estimation of lithium-ion batteries in electric vehicles. *Energy Procedia* 2016; 88: 998-1004, <https://doi.org/10.1016/j.egypro.2016.06.125>.
61. Wang Y, Zhang C, Chen Z. An adaptive remaining energy prediction approach for lithium-ion batteries in electric vehicles. *Journal of Power Sources* 2016; 305: 80-88, <https://doi.org/10.1016/j.jpowsour.2015.11.087>.
62. Wei P, Lu Z, Song J. Variable importance analysis: A comprehensive review. *Reliability Engineering and System Safety* 2015; 142: 399-432, <https://doi.org/10.1016/j.res.2015.05.018>.
63. Weng C, Cui Y, Sun J, Peng H. On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression. *Journal of Power Sources* 2013; 235: 36-44, <https://doi.org/10.1016/j.jpowsour.2013.02.012>.
64. Werner D, Paarmann S, Wetzel T. Calendar aging of li-ion cells-experimental investigation and empirical correlation. *Batteries* 2021, <https://doi.org/10.3390/batteries7020028>.
65. Wik T, Fridholm B, Kuusisto H. Implementation and robustness of an analytically based battery state of power. *Journal of Power Sources* 2015; 287: 448-457, <https://doi.org/10.1016/j.jpowsour.2015.03.165>.

66. Xiong R, Zhang Y, He H et al. A double-scale, particle-filtering, energy state prediction algorithm for lithium-ion batteries. *IEEE Transactions on Industrial Electronics* 2017; 65(2): 1526-1538, <https://doi.org/10.1109/TIE.2017.2733475>.
67. Xiong R, Zhang Y, Wang J et al. Lithium-Ion Battery Health Prognosis Based on a Real Battery Management System Used in Electric Vehicles. *IEEE Transactions on Vehicular Technology* 2019; 68(5): 4110-4121, <https://doi.org/10.1109/TVT.2018.2864688>.
68. Zhang X, Wang Y, Liu C, Chen Z. A novel approach of remaining discharge energy prediction for large format lithium-ion battery pack. *Journal of Power Sources* 2017; 343: 216-225, <https://doi.org/10.1016/j.jpowsour.2017.01.054>.
69. Zhang Y, Tang Q, Zhang Y et al. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nature Communications* 2020; 11(1): 6-11, <https://doi.org/10.1038/s41467-017-02251-3>.