Applied Soft Computing 111 (2021) 107670

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Comparing seven methods for state-of-health time series prediction for the lithium-ion battery packs of forklifts

Matti Huotari [a,*], Shashank Arora [b], Avleen Malhi [a,d], Kary Främling [a,c]

[a] Department of Computer Science, Aalto University, Espoo, Finland
[b] Department of Mechanical Engineering, Aalto University, Espoo, Finland
[c] Department of Computer Science, Umeå University, Umeå, Sweden
[d] Department of Computing and Informatics, Bournemouth University, Bournemouth, UK

ARTICLE INFO

Article history:
Received 5 December 2020
Received in revised form 21 June 2021
Accepted 23 June 2021
Available online 1 July 2021

Keywords:
Electrical vehicles
State-of-health for lithium-ion batteries
Machine learning
Neural networks
Timeseries prediction

ABSTRACT

A key aspect for the forklifts is the state-of-health (SoH) assessment to ensure the safety and the reliability of uninterrupted power source. Forecasting the battery SoH well is imperative to enable preventive maintenance and hence to reduce the costs. This paper demonstrates the capabilities of gradient boosting regression for predicting the SoH timeseries under circumstances when there is little prior information available about the batteries. We compared the gradient boosting method with light gradient boosting, extra trees, extreme gradient boosting, random forests, long short-term memory networks and with combined convolutional neural network and long short-term memory networks methods. We used multiple predictors and lagged target signal decomposition results as additional predictors and compared the yielded prediction results with different sets of predictors for each method. For this work, we are in possession of a unique data set of 45 lithium-ion battery packs with large variation in the data. The best model that we derived was validated by a novel walk-forward algorithm that also calculates point-wise confidence intervals for the predictions; we yielded reasonable predictions and confidence intervals for the predictions. Furthermore, we verified this model against five other lithium-ion battery packs; the best model generalised to greater extent to this set of battery packs. The results about the final model suggest that we were able to enhance the results in respect to previously developed models. Moreover, we further validated the model for extracting cycle counts presented in our previous work with data from new forklifts; their battery packs completed around 3000 cycles in a 10-year service period, which corresponds to the cycle life for commercial Nickel–Cobalt–Manganese (NMC) cells.

1. Introduction

Efficient transportation systems can improve the flow of goods and diminish the amount of energy used. In this regard, electric vehicles (EVs) have gained attention, especially in the area of forklifts, that now are not only energy-efficient but also safer for the drivers as they produce no fumes, vibrate less and are quieter than combustion-engine-powered forklift trucks. Lithium-ion batteries are widely employed to power these EVs. A battery's ability to store and deliver electrical energy is expressed by a measure, battery state of health (SoH), that is used for monitoring and for controlling these batteries in order to maximise their availability for operation. Heat generation in battery packs increases as they age. If unchecked, it can cause internal short circuits and compromise the safety of EV passengers and first responders. Therefore, battery packs in EVs are replaced when their SoH decreases below 80% [1]. However, these lithium-ion batteries can still be used for less-demanding grid-connected energy storage applications, i.e., the batteries can have a second-life use as components of an energy storing system for the sustainable energy management of a Smart City [2–4].

Overall, making battery related forecasts remains difficult in such manner that it generalises well [5]. For data-based approaches, there is a trade-off between complex hypotheses that fit the training data well, and simpler hypotheses that may generalise better [6,7]. For this paper, we are in possession of a unique data set of 45 lithium-ion battery packs with large variation in the data [8]. We compared five regression methods for making predictions for the SoH timeseries: gradient boosting (GB) [9], light gradient boosting (LGB) [10], random forests (RF) [9], randomised extra trees (ETR) [11], extreme gradient boosting (XGB) [12]. We also compared two other methods: long short-term memory networks (LSTM) [13] and long short-term memory network and convolutional neural network (LSTM–CNN) method [14]. Based

* Corresponding author.
E-mail address: matti.huotari@aalto.fi (M. Huotari).

https://doi.org/10.1016/j.asoc.2021.107670

on these methods, several models were grid searched to yield the best hyperparameters and the best model for each of the methods. Then the models yielded were back-tested in walk-forward manner [15]. We created a novel walk-forward algorithm for the data set that was re-framed as a supervised learning problem, that can utilise less steps and is therefore computationally lighter than the one-step-ahead walk-forward algorithm [16,17]; furthermore, the novel algorithm can be utilised to calculate the point-wise confidence intervals for the predictions. With this novel algorithm, we examined several prediction-period related parameters to detect the best time-span to utilise in order to yield reliable predictions. Moreover, as there is data available from several battery packs, we assessed the applicability of this data to the finally yielded best model, and then we verified the final model against five similar lithium-ion battery packs.

In the selection of the machine learning methods, we wanted to use methods that adapt well to timeseries regression, which, e.g., do not shuffle or split the data in such manner that the time-series sequence is broken. Secondly, we aimed at using a method with a fairly good record of speed and model performance in the field of machine learning, such as XGB [12]. Thirdly, we applied empirical mode decomposition to the SoH signal to determine if the decomposition data enhances the model performance [18]. Little research exists related to battery packs' state-of-health predictions; therefore, we have identified a few research gaps, based on which we have formulated the following research questions:

1. How can we obtain a robust SoH estimation model for lithium-ion battery packs with low error margins?
2. How can we validate the correctness of the used model?

Contributions

This paper contributes to the literature by introducing a novel GB model for SoH prediction based on real-world application of lithium-ion battery packs in forklifts and implementation of a novel walk-forward algorithm [9,19] for validating the models. The implementation of the novel walk-forward algorithm was first tested against a public data set on household power consumption [20]. Moreover, we further validated the model for extracting cycle counts presented in our previous work [21] with data from new forklifts; their battery packs completed around 3000 cycles in a 10-year service period, which corresponds to the cycle life for commercial Nickel–Cobalt–Manganese (NMC) cells published in [1].

Organisation

The remainder of the paper proceeds as follows. First, the related literature in presented Section 2 and the materials and methods in Section 3. The model development steps are described in Section 4. Subsequently, the main findings are presented in Section 5, followed by the discussion in Section 6. The conclusions and future work are discussed in Section 7.

2. Literature review

The lithium-ion battery packs in EVs and grid storage systems can benefit from the added reliability and safety assurance provided by a fast, yet accurate, SoH prediction. Traditionally, SoH forecasting has relied on equivalent-circuit models; however, more recently statistical and machine-learning techniques have been proposed, including ARIMA based statistical methods [22–24], neural networks [22,25,26], Gaussian processes [27,28], support vector machines [29–31], ensemble machine learning methods [32,33] and the XGB [22,34]. The success of the studies cited above demonstrates the capabilities of these approaches. Nevertheless, it is also known that the relationship between the basic signals and the SoH is complex under real conditions [35,36]. In addition, these studies focus on single cells, although the voltages available through the state-of-the-art single cells are insufficient for supporting an electric driveline. Therefore, many cells needs to be combined in series and in parallel to build up battery packs that are then used as energy sources in the EVs. Due to the manufacturing inconsistencies and differences in working environments, the behavior of each battery pack varies in real-life applications. Furthermore, this variation will become even larger as the cells age. As a result, an estimation based on the unit cell model will be inaccurate for real-world applications [37].

Therefore, in this study, in contrast to several recent studies that model a battery pack based on unit cells (e.g., [38–40]), we employed a battery data set from lithium-ion battery packs that are used in electric forklifts [8]. The data was obtained from one of the companies in the industry. This data is unique to our knowledge, as we have not found data repositories related to the forklift battery data, except on the cell level [41].

Table 1
Summary of the basic signals.

Basic signals
Time stamp of the data, 1 min interval
Measured voltage V
Measured current A
SOC %
Ambient temperature °C

3. Materials and methods

3.1. Data set description

For this study, the data set consists of 45 three-year timeseries that are derived from lithium-ion battery packs used in electric forklifts. The nominal capacity of the battery packs is 220 kW. We were told that the data was collected from different countries and continents. From the data we observed that the mean monthly ambient temperature for the forklifts ranged from 21.5 °C to 32.3 °C with instantaneous temperatures beyond this range. The data was collected using sensors selected by the battery manufacturer that were attached to the batteries. The raw data was sent to a local hub according to the date of the data collection and the serial number of the battery. The selected data points, comprised of current, voltage, and the ambient temperature, were taken every minute (Table 1). Based on these data points, we extracted more features using our feature extraction method [21].

In this study, the number of timeseries available was greater than in our previous study [21]; however, the new data supported our previous findings. We observed that the median number of occurrences of the charging pulses was two in a day; here a charging pulse is defined as a period between 5–30 min when the battery's state-of-charge increases. However, there were days when there were no charging pulses (timeseries are irregular). For more information on the basic and on the derived signals, see our previous paper [21].

3.2. Proposed methodology

Although there are several statistical methods [42] or neural network based methods [43] to compose a solution for incomplete or irregular timeseries, in this paper we wanted to utilise regression methods that require regular timeseries. Furthermore, we wanted to compare the results with our previous ARIMA based results with results that are directly comparable [21]. For these reasons, the irregular timeseries based

on one-minute measurements was aggregated to regular daily timeseries; the number of resulting timesteps was around 1000 for the battery packs.

The overall prediction target, state-of-health, is defined as the ratio of current capacity to the initial capacity of a battery; the SoH is denoted as timeseries in this paper [21]:

$$SoH(t) = \frac{C_n(t)}{C_0(t_0)} \tag{1}$$

The empirical mode decomposition is a mathematical time domain decomposition method, which can convert a group of timeseries into locally narrow band components, the intrinsic mode functions [44]. This method is applied to, e.g., asserting power quality [45], or predicting remaining useful lifetime of lithium-ion batteries [22].

A timeseries can be transformed by using an empirical mode decomposition, which in this case is denoted as:

$$SoH(t) = \sum_{i=0}^{N} c_i(t) + r_N(t) \tag{2}$$

where $c_i(t)$ are the intrinsic mode functions (IMFs) separated by instantaneous frequencies, $r_N(t)$ is the residue and N is the finite amount of decompositions obtained [44]. In this paper, for the needs of the model, the residue is used as the trend [46], although there are more refined trend extraction methods available [47].

The IMFs can be transformed by using the Huang–Hilbert transform. This is used to obtain the analytic signal, which can be presented in polar form [48], neglecting the residue. In this case this yields:

$$SoH(t) = \sum_{j=1}^{n} a_j(t) e^{i(\theta)_j(t)} \tag{3}$$

where $a_j(t)$ are the analytic signals [49]. The synthesised signal models a non-stationary and non-linear system analytically, i.e.,

$$SoH(t) = a(t)cos\theta(t) \tag{4}$$

where $a(t)$ is the instantaneous amplitude and $\theta(t)$ the instantaneous phase. The optimal values for the parameters of the synthesised signal models are tuned according to the signal in question; i.e, the IMFs and residue change as the data set changes, and decomposing parameters needs to be set accordingly to yield an accurate decomposition [18]. Finally, the instantaneous frequency is obtained through the derivative of the instantaneous phase, i.e.,

$$f(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} \tag{5}$$

The derived $f(t)$ is a tool for analyzing transient signals, such as battery pack SoH, whose constituent frequencies may change over time [50,51]. The lagged instantaneous frequency, $f(t-1)$, was used as one of the predictors for our models.

3.3. Selected SoH prediction methods

Extreme gradient boosting (XGB) is an ensemble of gradient boosted decision trees algorithm [9,52]. It uses decision trees where new trees improve the model consisting of those trees that are already part of the model. It is used, for example, for forecasting energy load [53] and for forecasting the battery cell state-of-charge (SoC) as represented in [54]. As for the other methods utilised, gradient boosting (GB) is described in [9], extremely randomised trees (ETR) in [11], random forests (RF) in [9], long short-term memory networks (LSTM) e.g. in [10] and convolutional neural networks (CNN) in [14].

3.4. Data preprocessing methods and performance metrics

The outliers in the initial data for each battery pack were eliminated by the interquartile range (IQR) method [55]. After removing the outliers, we imputed some missing daily data. The missing values were the mean of the daily values above and below the missing values.

After creating the predictors and targets for the models, we re-framed the multivariate timeseries as a supervised learning problem [9] in order to define the number of past time steps used for making a forecast and to define the number of prediction timesteps for the prediction horizon. In the model tuning phase, we split this re-framed data to training data and test data (more details on splitting methods and on numerical values utilised for the test and train sets is in the next section in Table 7). We evaluated the used prediction methods using four different metrics: the root-mean-squared error (RMSE) [56], the mean absolute error (MAE) [57], coefficient of determination, R^2 [58] and the explained variance (EVAR) as in Eq. (6). In this paper, the MAE and the RMSE are related to the SoH range 0–100(+) %. E.g., a MAE 1 implies that, on average, the forecast's distance from the true SoH value is 1. For this data set, a MAE value of 1 is significant as, e.g., the yearly degradation of SoH for battery (a) is around 2.2 percentage points [21].

$$\text{EVAR} = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}} \tag{6}$$

For EVAR and R^2 evaluation methods, the best possible score is 1.0; a baseline model predicting the mean ($\bar{y}$) has score 0; models with less skill than the baseline model will have negative scores.

3.5. Validation of models and calculation of the point-wise confidence intervals for the estimates by a novel walk-forward algorithm

We executed a basic comparison of models with walk-forward method to predict and to find the best model in terms of MAE [15,16]. A basic walk-forward method, that utilises expanding window and proceeds one-step-ahead at each iteration round, can be utilised for all comparisons [17]. Nevertheless, we wanted to find an approach that utilise computational resources sparingly as the timeseries can grow long (10-years or more) or there can be several models to be evaluated at the same time for a fleet of forklifts. The following parameters can be set for the novel algorithm: sample size (number of the latest observation windows used by the algorithm) and roll size (number of windows stepped over in an iteration) 1; these functionalities are not part of the standard machine learning (sklearn) library timeseries split for a multivariate data set re-framed as supervised learning problem in a simple manner. For the implementation of the algorithm, we used Pandas's append and del functions and sklearn's regression methods (GB, RF, XGB and ETR) [59]. It is noteworthy that the window size (number of past observations and future observation in the prediction window) was set in the model tuning phase; furthermore, each window size requires a model of its own [60].

As the algorithm utilises the sliding window method, the successive training sets are not super-sets of those coming before them, and this yields models that have more variation; however, at the same time, some of the training data is lost.

Moreover, the algorithm yields point-wise confidence intervals (CI) that quantify the uncertainty for the predictions [16,61]. In this paper, we added upper and lower confidence intervals (CI) to each of the point-wise predictions for the selected final model (Fig. 7). The standard error (SE_{SoH}) that is needed for yielding a confidence interval was calculated as:

$$\widehat{SE_{SoH}}(\hat{\mu}_t(n)) = \frac{\hat{\sigma}_t(n)}{\sqrt{n}} \tag{7}$$

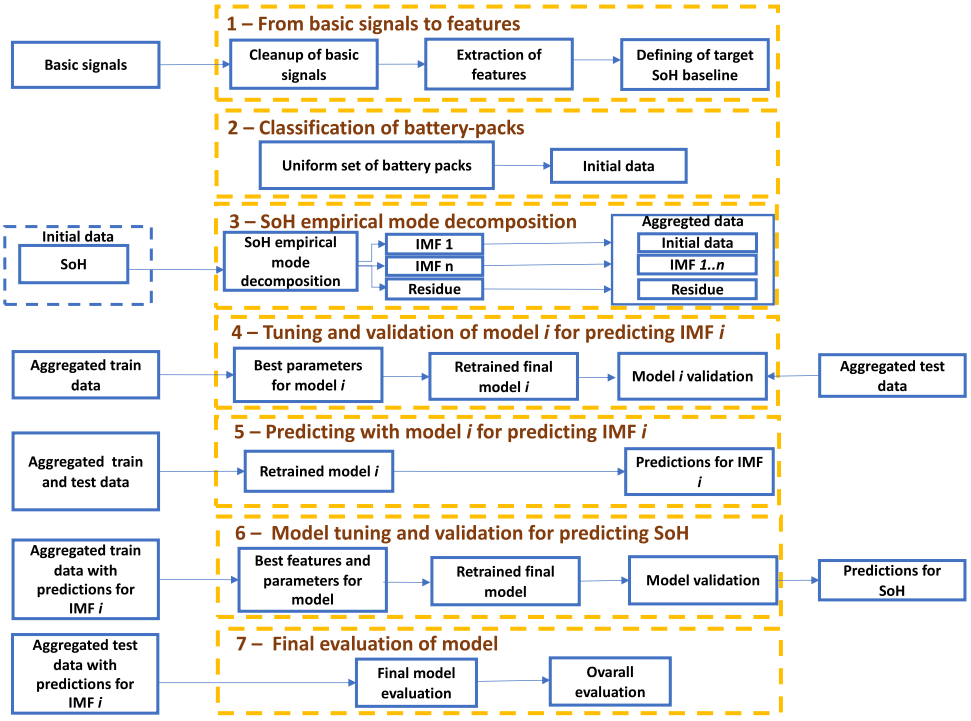

Fig. 1. The selected SoH prediction model development steps 1–2: data preparation. Steps 3–5: SoH decomposition and decomposition related predictions. Steps 6–7: the final model predicting the SoH with the model evaluation.

Algorithm 1 Walk-forward: sample predictions with point-wise confidence intervals

inputs:
  Obs: timeseries' observations that are re-framed as supervised learning problem (re-framed to windows)
  n_S: sample of windows utilised for testing
  n_{roll}: number of windows stepped over in a window roll
local variables:
  Te: a sample of timeseries' windows
  Tr: timeseries' windows preceding *Te*
  RTe: a rolled sample of timeseries' windows
  W_p: consequent predictors utilised for making predictions
outputs:
  $\widehat{SoH}$: SoH predictions for the sample
  CI: point-wise confidence interval for these SoH predictions
  W_t: consequent targets (ground truth)
require:
  $|Obs| > 1$
  $n_S > 0$

Tr,*Te* ← split *Obs* to train and test sets according to n_S
RTe ← [ ]
for $R \leftarrow 0$ to $|Te|$ **mod** n_{roll} **do**
  RTe ← append window (*Te*[*R*])
end for
$\widehat{SoH}$,*CI* ← [ ], [ ]
for $T \leftarrow 0$ to $|RTe|$ **do**
  W_p, W_t ← *RTe*[*T*] separate predictors and targets for this iteration step
  $\widehat{SoH}$ ← fit the model with *Tr* and predict with W_p
  CI ← append $\pm 1.98^{*}\mathbf{SE}(\widehat{SoH})$ ▷ Eq. (8)
  Tr ← append windows from *Te* until and including window *RTe*[*T*]
  if sliding window **then**
    Tr ← delete $|n_{roll}|$ windows from head of *Tr*
  end if
end for
return $\widehat{SoH}$, W_t, *CI*

where *n* are the prediction made by the novel Algorithm 1.

We used the 95% significance level for the point-wise confidence intervals, which corresponds to the Gaussian distribution critical value 1.96. Hence, a confidence interval was calculated as:

$$CI = 1.96\widehat{SE_{\mathrm{SoH}}}(\hat{\mu}_t(n)) \tag{8}$$

for a point-wise SoH prediction (Algorithm 1 above) [16,61].

4. Development of models

The overall flow of the model development is depicted in Fig. 1. In the steps 1–2, we cleaned the basic signals and extracted new ones [21]. Initially, we had to eliminate the adverse effect of severe transient failures, where, e.g., a sensor had sent erratic values. Furthermore, for selecting the battery packs for the model development, we scrutinised the ambient temperatures. On one hand, ambient temperature below zero may have had an adverse effect on the SoH of the battery [62]. On the other hand, relatively high ambient temperatures (> 32 °C) also have an adverse effect on the SoH [1,63,64]. Moreover, for the batteries used in forklifts, the ambient temperature fluctuation showed some seasonality for all battery packs; an example of this is in Fig. 3. For these reasons, for the model development, we used a set of battery packs with the same 32-month data record and the mean ambient temperature between the range mentioned above.

For the remaining battery packs, the following physical quantities were extracted: charging time, charging voltage, charging current and difference in state-of-charge (SOC) during chargings. From these we derived charging energy and charging cycles as described in [21].

For verifying that the cycle count calculation method developed in the previous paper was valid for the new battery packs available for the study this time, we selected batteries' (a–f) timeseries randomly from the suite of 45 batteries for making graphs and for making initial reasoning based on those graphs; these new results supported the findings of our previous study. In the monthly averaged SoH data, there were some obvious outliers; moreover, there was remarkable fluctuation day-to-day. Overall, the SoH values for a typical battery showed a linearly fitted trend downwards. However, small increases in the capacity after a slow cycle or a rest period may result in the SoH exceeding 100% [32]. This can be visually confirmed from the plots for the selected six batteries in Fig. 2. We decided to use the battery (a) in the further development of the model in more detail; however all of the six batteries are taken into account in the final verification.

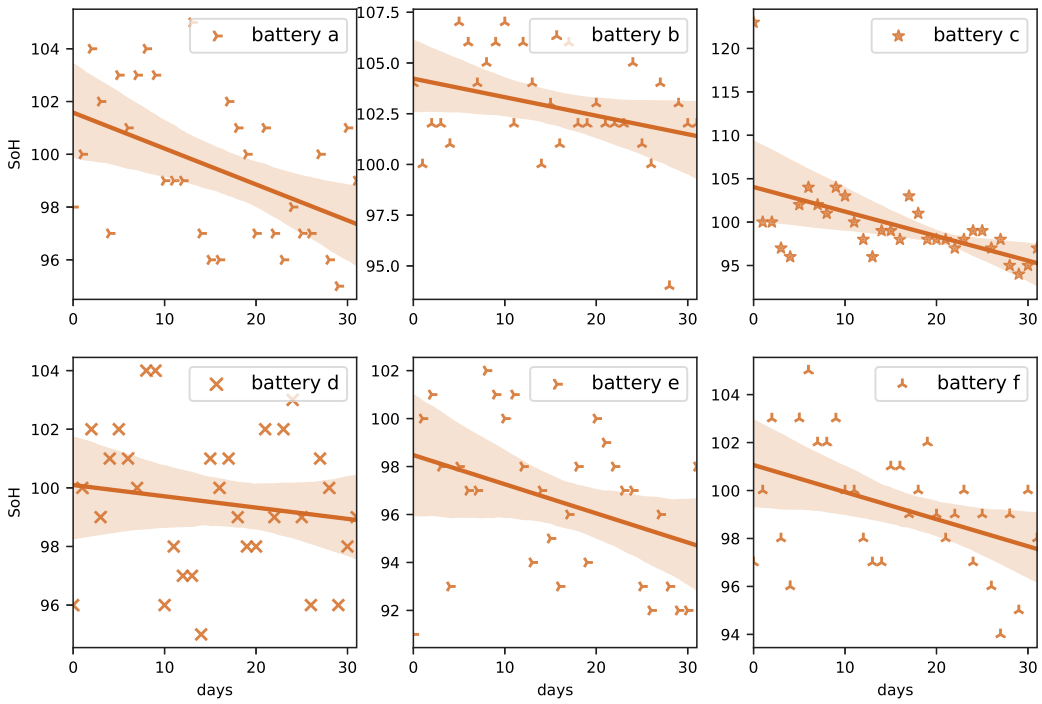

Fig. 2. State-of-health trends for batteries (a) – (f).

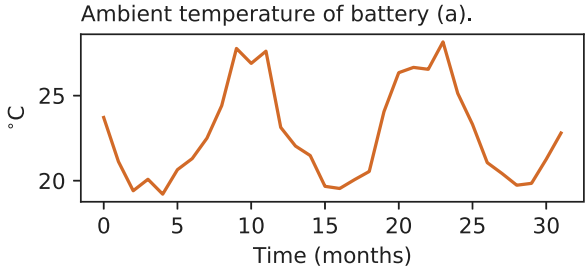

Fig. 3. Ambient temperature for battery (a).

From this initial data the outliers were detected and removed by the interquartile range selection method; any observations that were more than 1.5 × IQR below the Q1 or more than 1.5 × IQR above the Q3 were considered as outliers [55]; some 3%–5% of the data was discarded. Then the missing daily values were imputed by taking the average of the values before and after the missing values. As the next step for the model generation, the step 3 in Fig. 1, we selected the complete empirical mode decomposition (CEEMDAN) implementation [18], which fundamentally implements the method as described in the methodology Section 3.2 (an example of decomposition's results is in Fig. 4). To obtain the instantaneous frequency for the SoH, we performed the Huang–Hilbert transform at this stage (example in Fig. 9), lagged it by one time step (t-1), and added it as a predictor. The summary of the features after the feature extraction is in Table 2.

For the IMF and residue predictions, in the model tuning and validation phases (steps 4 – 5 in Fig. 1), 70% of the data set was used for tuning and 30% for validation. We selected XGB randomly as the method to use, albeit proven one in the field of batteries [22,34]. For the tuning, we used the predictors presented in Table 2 and tuned the model using 10-fold timeseries cross-validation; the tuning results are presented in Table 3. The best models were verified with the test set. Finally, the resulted predictions for each of the IMFs and for the residue were added to the final data set as additional predictors maintaining the train and test split for avoiding data leakage.

Table 2
Selected features used by a model. The ticks represent the features used by a model.

Signals/Target (predictors)	IMFs	SoH (basic)	SoH (basic+IMFs)
Time stamp of the data	✓	✓	✓
Voltage V	✓	✓	✓
Current A	✓	✓	✓
SOC %	✓	✓	✓
Ambient temperature °C	✓	✓	✓
Charging length in minutes	✓	✓	✓
Energy Wh	✓	✓	✓
Voltage difference ΔV	✓	✓	✓
Cycle	✓	✓	✓
(t-1) lagged instantaneous frequency	✗	✓	✓
Prediction for residue	N/A	✗	✓
Prediction for IMFs	N/A	✗	✓

4.1. Model tuning and initial model comparison for predicting the SoH of a battery pack

Next, we generated the models for predicting SoH (steps 6–7 in Fig. 1). We selected seven different up-to-date methods to ensure that we find a good model for the prediction. We tuned the GB, LGB, RF, ETR and XGB models with the predictors presented in Table 2. For this purpose, we re-framed the timeseries to windows with size seven (six past observations and prediction horizon of one) and utilised 10-fold timeseries cross-validation in the grid search of the best hyperparameters. For each of the methods, we tuned two models with SoH as the target: the first model utilised basic predictors and the second utilised basic predictors together with intrinsic mode function (IMF) and residue prediction values (i.e., $\tilde{SoH}(t)$). The summary of the tuned hyperparameters is provided in Table 4.

Furthermore, two simple LSTM and CNN–LSTM models were grid searched (network structure in Table 5) for finding the

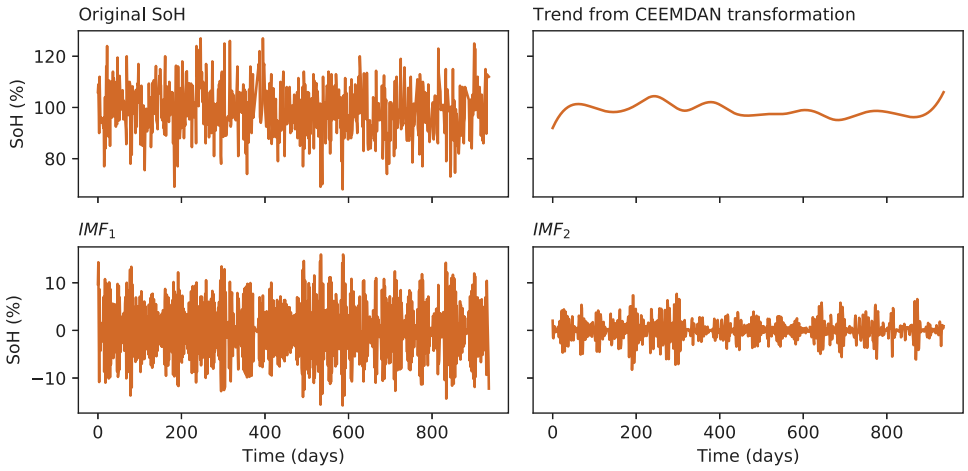

Fig. 4. SoH, trend and two first IMFs obtained from complete empirical mode decomposition transformation for battery (a).

Table 3
Overview of the selected models for predicting intrinsic mode function and residue values.

Target		IMF1	IMF2	IMF3	Residue
Method		XGB	XGB	XGB	XGB
Number of estimators		1000	500	250	250
Maximum depth		2	2	2	3
Column samples	By level	0.8	0.8	0.8	0.8
	By node	0.8	0.5	0.5	0.8
	By tree	0.8	0.8	0.8	0.5

best hyperparameters for them. In the evaluation phase, we normalised the used data (each predictor variable and the target variable), and repeated each model evaluation 10 times and averaged the results after inverting the normalisation for the final results. All of the predictors were used for these models (Table 2). This grid search method was used as an initial attempt to yield a network structure for this data set as no prefitted neural network model exists for this data.

For the five methods with the best EVAR yielded (in Table 7), we expanded the number of models. We utilised the same methods but increased the number of window sizes utilised. For each new window size, we tuned a model of its own. After tuning, we made a comparison between all the yielded models with the novel algorithm 1 for finding the optimal values for samples sizes, rolling window sizes and roll step sizes for the models for this data set (Section 4.3).

4.2. Verification of the novel walk-forward algorithm and calculating the point-wise confidence intervals with an external data set

We verified the novel walk-forward method for yielding the point-wise confidence intervals against a public data set (the household power consumption data set [20]); this data was aggregated to monthly values. The model that was used to predict the monthly power consumption was a simple XGB model (number of estimators: 50, maximum depth: 2); it yielded the RMSE 0.08, which exceed the naïve model's RMSE 0.11. For the verification, the number of observation windows used for walk-forward was 30, the size of the sliding window was 7 and the number of rolled over windows was 4. The yielded prediction results with the corresponding point-wise confidence intervals are in Fig. 5. The CI does not fluctuate significantly nor show a clear trend; the model seems to be stabile [16]. As these results supported the theoretical basis [9] for its use as a verification method, the

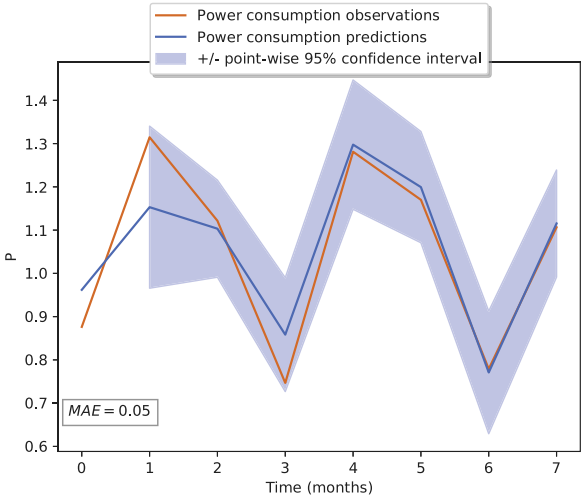

Fig. 5. The graph of the household power consumption predictions with the XGB model and the point-wise CI for the predictions. Notice that the confidence interval varies due to, e.g., the used aggregation method, the rolling window size and the underlying model's stability.

point-wise confidence interval calculation method was used in our model development as well.

4.3. Verification of the novel walk-forward method and calculating the point-wise confidence intervals with battery data set

For the finally selected methods (in Section 4.1), we used the predictors presented in Table 2 and tuned the models using 10-fold timeseries cross-validation; the tuning results are presented in Table 6. After the tuning, the yielded models were verified with the novel algorithm with both expanding and sliding windows. As a summary, the following sizes were utilised for samples: 14, 30 and 90 days, for windows: 7, 14 and 30 days (with prediction horizon 1 included in these figures), and for window rolls: 1, 2, 7, and 14 days. In order to make enough repetitions, the algorithm parameters were set so that sample size $>$ window size $>$ 2 x roll size, except for one case where roll and window sizes were set to be equal. It can be noted that sample size determines the initial test set, from which walk-forward starts to roll the window towards the newest observations in the test set.

Table 4
Overview of the selected regression models for predicting SoH (window size 7).

Target		SoH	SoH	SoH	SoH	SoH	SoH	SoH	SoH	SoH	SoH
Method		GB	EMD-GB	LBG	EMD-LGB	RF	EMD-RF	XGB	EMD-XBG	ETR	EMD-ETR
Number of estimators		1000	1000	250	500	500	1000	1000	1000	500	1000
Maximum depth		4	4	8	6	6	5	2	2	6	6
Column samples	By level	N/A	N/A	N/A	N/A	N/A	N/A	0.8	0.8	N/A	N/A
	By node	N/A	N/A	N/A	N/A	N/A	N/A	0.8	0.8	N/A	N/A
	By tree	N/A	N/A	0.8	0.8	N/A	N/A	0.8	0.8	N/A	N/A
Subsamples		0.8	0.8	0.8	0.8	N/A	N/A	1	1	N/A	N/A
Subsample frequency		N/A	N/A	5	5	N/A	N/A	N/A	N/A	N/A	N/A

Table 5
Overview of the LSTM and CNN–LSTM models for predicting SoH (window size 7).

LSTM			CNN–LSTM		
Layer (type)	Output shape	Number of parameters	Layer (type)	Output shape	Number of parameters
lstm_10 (LSTM)	(None, 200)	172 800	conv1d_2 (Conv1D)	(None, 12, 64)	2944
repeat_vector_5 (RepeatVecto)	(None, 7, 200)	0	conv1d_3 (Conv1D)	(None, 10, 64)	12 352
lstm_1 (LSTM)	(None, 7, 200)	416 800	max_pooling1d_1 (MaxPooling1)	(None, 5, 64)	0
time_distributed_2 (TimeDist)	(None, 7, 50)	10 050	flatten_1 (Flatten)	(None, 320)	0
time_distributed_3 (TimeDist)	(None,7,1)	51	repeat_vector_1 (RepeatVecto)	(None, 7, 320)	0
			lstm_1 (LSTM)	(None, 7, 200)	416 800
			time_distributed_2 (TimeDist)	(None, 7, 50)	10 050
			time_distributed_3 (TimeDist)	(None,7,1)	51

Table 6
Overview of the further selected regression models for predicting SoH (window sizes 14–90).

Target			SoH	SoH	SoH	SoH	SoH
Method			GB	EMD-GB	RF	EMD-RF	EMD-ETR
Window	14	Number of estimators	250	250	250	1000	1000
		Maximum depth	4	4	5	6	6
		Subsamples	0.8	0.8	N/A	N/A	N/A
Window	30	Number of estimators	1000	1000	500	500	500
		Maximum depth	5	5	5	6	6
		Subsamples	0.8	0.8	N/A	N/A	N/A
Window	90	Number of estimators	1000	1000	250	250	500
		Maximum depth	4	4	5	5	6
		Subsamples	0.8	0.8	N/A	N/A	N/A

Table 7
Summary of the model comparisons (EVAR is the explained variance). The five best grid searched models are in boldface.

Battery	Model	Samples	EVAR	MAE	RMSE
Battery (a)	**GB**	**70%–99% of data**	**99.9**	**0.21**	**0.28**
Battery (a)	**EMD-GB**	**70%–99% of data**	**99.9**	**0.20**	**0.28**
Battery (a)	**RF**	**70%–99% of data**	**99.9**	**0.22**	**0.31**
Battery (a)	**EMD-RF**	**70%–99% of data**	**99.9**	**0.24**	**0.32**
Battery (a)	LGB	70%–99% of data	99.2	0.48	0.81
Battery (a)	EMD-LGB	70%–99% of data	99.3	0.45	0.75
Battery (a)	EXT	70%–99% of data	99.6	0.51	0.59
Battery (a)	**EMD-EXT**	**70%–99% of data**	**99.7**	**0.43**	**0.53**
Battery (a)	XGB	70%–99% of data	99.3	0.53	0.76
Battery (a)	EMD-XGB	70%–99% of data	99.6	0.44	0.59
Battery (a)	EMD-LSTM	70% of data	–	7.70	9.47
Battery (a)	EMD-CNN LSTM	70% of data	–	8.09	10.27

5. Results and discussion

5.1. Results for battery pack (a) and comparison with the previous methods

The best initial model for the battery (a) was yielded by gradient boosting with additional predictors (i.e., SõH(t)) and with window size 7. (See (EMD-GB) and the rest of the results in Table 7). This model yielded mean absolute error 0.20 and RMSE of 0.28 that are in the low error ranges. Furthermore, it can be noted that all of the four best regression models were close-by each other in terms of MAE. For the LSTM and LSTM–CNN it can be noted that the results indicate that an optimal neural network structure and parameters were not found this time.

In the further validation, we tuned new models, and then we back-tested them with the novel walk-forward method using the battery (a). The results (two of the best scores and the worst score for each model) are presented in Table 8. The best initial model for the battery (a) was gradient boosting without the additional predictors (i.e., without SõH(t)) and with window size 14.

This model yielded mean absolute error 0.18 and RMSE of 0.20, and is the final best result for this paper. Furthermore, for gradient boosting and for random forests models, the overall difference between the best and the worst MAE for all of the submodels was narrow (< 0.1). The best models were yielded without the additional predictors (i.e., without SõH(t)). In contrast to these results, for the light gradient boosting, extreme boosting and extra trees, the overall difference between the best and the worst MAE for all of the submodels was typically wider than 0.1. The best models were yielded utilising the additional predictors (i.e., with SõH(t)). A conclusion is that the decomposed SõH(t) predictions used as predictors slightly enhanced some models' overall prediction accuracy. The expanding window method yields, in majority of the cases, slightly smaller MAE and RMSE values for this data set than the sliding window method. This is according to the general findings in the industry for data with relatively few samples [17]. The window roll sizes 1–7 yielded models that had the best and the worse MAE results close by each other with the same window

Table 8
The two best results and the worst result for SoH predictions for the models with $\tilde{SoH}(t)$ predictors (EMD-) and without them for battery (a). All models are verified with both expanding and sliding window. The best model yielded, is in boldface.

Roll type	*Expanding window*					*Sliding window*				
Model	Sample	Win	Roll	MAE	RMSE	Sample	Win	Roll	MAE	RMSE
	30	**14**	**1**	**0.18**	**0.23**	14	7	1	0.20	0.24
GB	30	14	2	0.19	0.23	30	7	1	0.20	0.25
	14	7	2	0.26	0.33	14	7	2	0.25	0.33
	30	14	2	0.19	0.23	30	14	1	0.20	0.25
EMD-GB	90	30	1	0.20	0.22	90	14	1	0.21	0.25
	14	7	2	0.26	0.33	90	30	7	0.25	0.32
	30	7	2	0.20	0.27	30	7	2	0.20	0.27
RF	30	7	1	0.22	0.27	30	7	1	0.22	0.28
	90	30	7	0.29	0.39	90	30	2	0.29	0.45
	30	14	1	0.21	0.27	30	14	2	0.20	0.27
EMD-RF	30	14	2	0.21	0.27	30	14	1	0.21	0.27
	90	30	7	0.30	0.42	14	7	2	0.29	0.33
	90	30	7	0.39	0.49	90	30	7	0.41	0.50
EMD-ETR	90	7	1	0.46	0.55	30	14	1	0.42	0.51
	14	7	2	0.54	0.63	14	7	2	0.57	0.66

Table 9
The GB model comparative analysis with previous ARIMA model results [21].

	GB model	ARIMA
RMSE	1.56	2.68
R2	0.91	−0.26

Table 10
Evaluation results of verifying the GB model with data from batteries b-e.

Battery	(b)	(c)	(d)	(e)	(f)
EVAR	99.9%	99.8%	98.8%	99.8%	99.9%
MAE	0.18	0.29	0.20	0.15	0.27
RMSE	0.26	0.39	0.90	0.27	0.30
R^2	1.0	1.0	0.99	1.0	1.0

size (e.g. 0–0.08 difference in MAE in Table 8), i.e., a roll size of 7 can be applied to this data set without affecting the MAE results. Moreover, we spot-tested some roll sizes that do not have overlap with the previous window (e.g. window of size 14 and window roll of size 14); however, the MAE deviated in random manner, and in many cases by 50% from the results with the same window size but with a smaller roll. This indicates that a window roll that is up to 23%–28% of the utilised window size is applicable to this data set for yielding reliable results.

The best model (GB) was refitted to battery (a), after which we made predictions anew (Fig. 8). We yielded the MAE loss function value of 1.52 and the goodness of the fit, R^2, 0.91; the model outperformed the ARIMA model that we introduced in our previous paper (Table 9), although neither of these models perform well over the entire 32-month period.

Furthermore, as this timeseries model development setup was designed to predict SoH for the near future (sample sizes used were the newest 14–90 days and window sizes were 7–30 days), we refitted the model to the 3 nearest months (Fig. 7). We yielded the MAE loss function value of 0.21 and the goodness of the fit, R^2, 1, which indicates that the model predicts well over three-month-period. Moreover, it should be noted that the model can be overconfident in its predictions indicated by a R^2 score that is one. Therefore, in order to set confidence intervals for estimating model stability, the novel walk-forward method was applied to yield point-wise confidence intervals. (See Section 5.3).

5.2. Results for a set of batteries

For testing the timeseries of all batteries, we evaluated them with the Wilcoxon signed-rank test [65,66]. The test results were used to assess if a uniform forecast model can be applied to the set of batteries or not. Our hypothesis was that we can apply our model to the set of battery packs, as they come from the same factory, have the same calendar age and are used in the similar forklifts. The hypothesis (H0) was set as follows: the sample distributions from different batteries were related to the battery (a). Wilcoxon yielded that 65% of the batteries had the same distribution as battery (a) (failed to reject H0), and consequently 35% had different distribution (rejected H0).

Amongst the batteries in the same distribution, the batteries (b–f) were scrutinised in more detailed manner. For the verification, the best model extracted for battery (a) was applied to the five battery packs. The verification yielded MAE between 0.15–0.29 and goodness of the fit, R^2, between 0.99–1.00 (Table 10 and Fig. 6). The overall evaluation results for the set of batteries are promising; however, there is a need for further analysis on, e.g., the environmental factors and the length of the battery data on the battery SoH forecast; these may make a difference for enhancing the model and its reliability.

5.3. Results for calculating the point-wise confidence intervals for battery (a)

We calculated the confidence intervals for the SoH predictions for battery (a) in order to evaluate the overall model behavior. As depicted in Figs. 7 and 8, we used sample size of 3 and 30 months; we refitted the best GB model to the data and depicted the results. For the 3-month sample size the point-wise confidence intervals varied so that, at its narrowest, the range 88.0–91.9 covered the true prediction with the 95% likelihood and, at it widest, the range 103.6–126.6 covered the true prediction with the 95% likelihood; overall, for a 3-month period this GB model yielded reasonable predictions.

As a final remark on the results, calculating the predictions with confidence intervals with data re-split and model refit at every walk-forward iteration step is computationally heavy; however for the relatively small number of observations (around 1000 for 3-year period and 3700 for the 10-year period for one battery pack), this is tolerable; furthermore the window roll diminishes the number of calculations.

5.4. Results for instantaneous frequency

As a second but last result for this paper, the analytic signal derived from the decomposed SoH provides means to detect

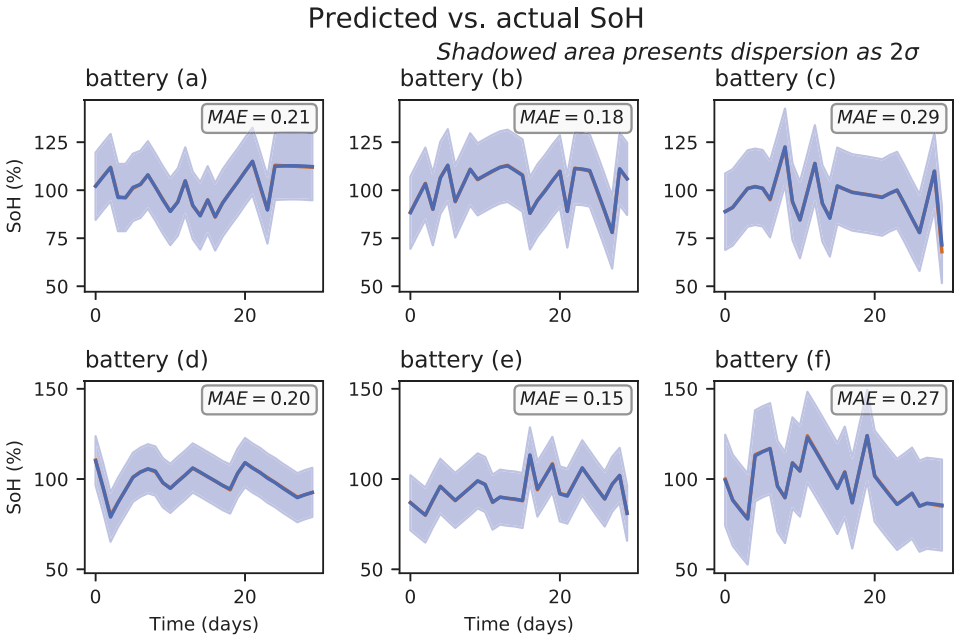

Fig. 6. Battery (b–f) GB predicted SoH values. The blue line is the predicted and the orange is the observed SoH. The shadowed area is +/− two standard deviations.

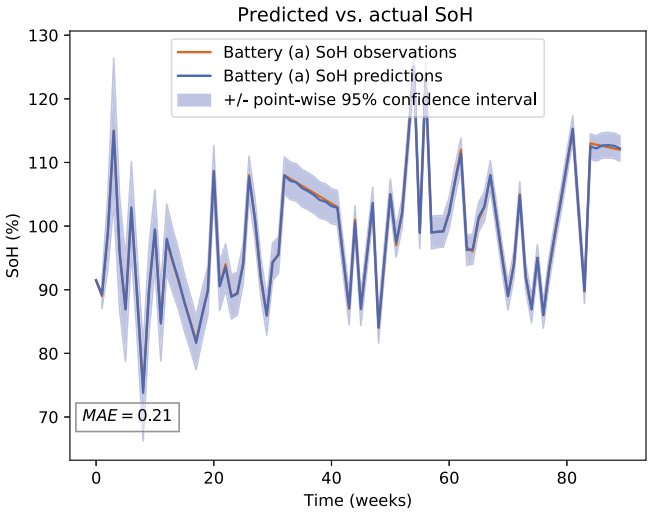

Fig. 7. Battery (a) EMD-LGB model predicted SoH values for three months (90 days) with confidence interval. The blue line is the predicted and the orange is the observed SoH. The shadowed area indicated point-wise 95% confidence interval for predictions.

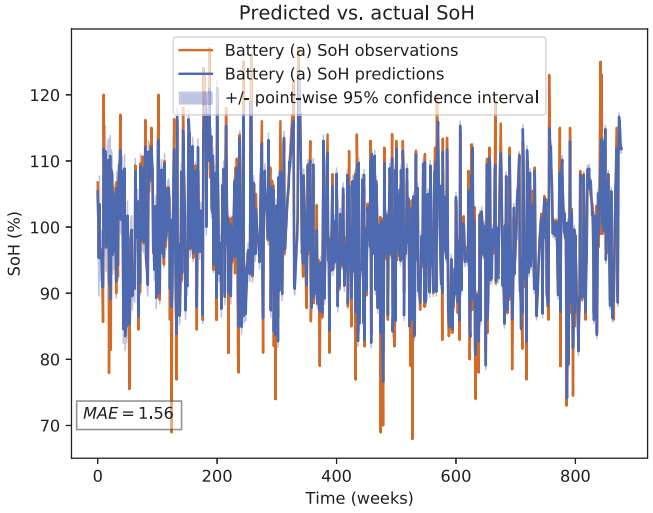

Fig. 8. Battery (a) SoH values for 30 months with the confidence interval; the blue line is the predicted and the orange is the observed SoH. The shadowed area indicates point-wise 95% confidence interval for predictions.

changes i.e., it provides means to analyze if the SoH starts to deteriorate, or, to change less frequently or more frequently than before; this information alleviates the decision to initiate an inspection of the battery in the field, or, to change the model. For example, the persistent drop in the instantaneous frequency (Eq. (5)) around the day 900 may indicate a general trend that may require revamping the model if this change is permanent (Fig. 9). Furthermore, there is a peak around day 380 in the figure. Inspecting the corresponding data file revealed that the SoH values fluctuated 30% between consequent days during several days; this indicated some kind of transient failure, and revealing its root cause would require a further inspection.

5.5. Results for the equivalent cycles

Extrapolating from the equivalent cycle count graph, if the usage behavior remained unchanged, the model estimated the truck's battery pack to complete around 3000 cycles in a 10-year service period (Fig. 10). The estimated cycle life corresponds to the published cycle life for the commercial Nickel–Cobalt–Manganese (NMC) cells [1], and support the findings in our previous study [21].

6. Discussion

In this paper, we have demonstrated the applicability of GB method for model development for battery state of health (SoH) predictions under circumstances when there is little prior information available about the batteries. It demonstrates how the

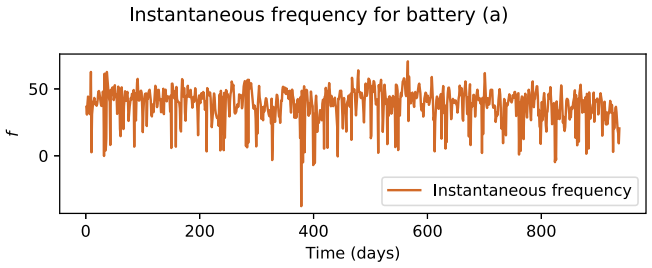

Fig. 9. The used Hilbert–Huang transformed instantaneous frequency for battery (a).

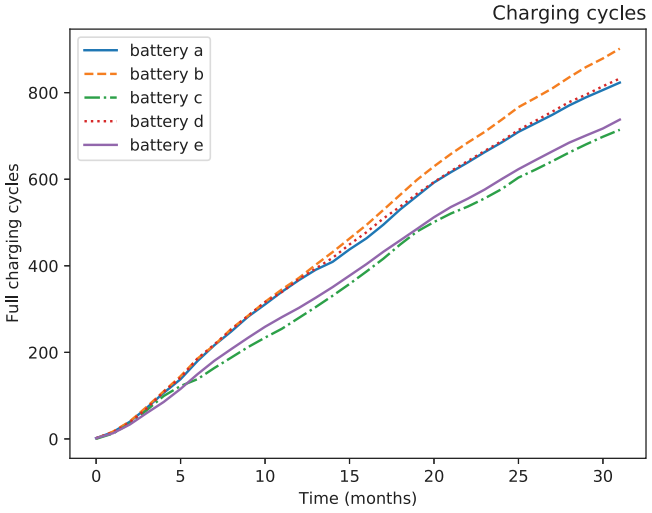

Fig. 10. The equivalent cycles for the batteries a, b, c, d and e.

supervised-learning-enabled SoH prognosis can effectively exploit the data from multiple cells in lithium-ion batteries from 45 EV forklifts and significantly improve the forecasting performance compared to previous models.

In the model development work we could verify the results from the previous study with new data; e.g., the cycle life-time produced values that matched well with the values published by the cell manufacturers. This indicated that we were fairly successful in the development of the basic features and extracting their values. Furthermore, the developed GB model predicts the SoH well with the loss function value of MAE_{SoH} 0.18 and with goodness of the fit, R^2 score 1 over a period of month.

Furthermore, we have shown that for a set of batteries, the Wilcoxon test yielded that the set of batteries come from the same distribution; this and the results using a common prediction model showed some promising loss function (MAE) and goodness of fit results for the set of batteries. However, it may well be that the different operating environments for the batteries result in SoH patterns that are not reliably enough captured by our GB model. Moreover, in the field of batteries, non-availability of more data sets with relatively long timeseries has is a known issue [67]. This was also the case with our unique set of data as the timeseries were relatively short (32 months), which in turn indicates a drawback in our models. For example, the battery timeseries may show seasonality in the long run that we were unable to capture.

With the GB mode, utilising the novel walk-forward algorithm for battery (a) over a period of three months, we yielded the MAE_{SoH} loss function value 0.21 including the 95% point-wise prediction confidence intervals. As for the environmental factors, the battery aging is known to be non-linear process, and the SoH deteriorates rapidly towards the end-of-life. For this reason, we introduced intrinsic frequency method to detect changes in the target SoH behavior for defining the point when the underlying GB model needs changes; the exact process how to implement this into the model needs further introspect. More data is needed to model end-of-lifetime behavior for the batteries, especially when taking into account that our model covered 3-year life-span out of around 10 years of expected life-time for the lithium-ion batteries for the EVs and for the 2nd life use. Although the model was developed with the data from EVs, it can be applied to the lithium-ion batteries that do not meet any longer the requirements of an EV application. It can still be used for the less-demanding grid-connected energy storage applications such as the battery energy storage system (BESS) for the sustainable energy management [68].

7. Conclusions & future work

In this paper, we have demonstrated the applicability of the gradient boosting model for predicting battery state of health (SoH) timeseries under circumstances when there is little prior information available about the batteries. It demonstrates how supervised-learning-framed SoH prognosis can effectively exploit data from multiple cells in lithium-ion batteries from 45 EV forklifts to significantly improve the forecasting performance. The GB model predicts SoH well with the loss function value of MAE_{SoH} = 0.21 and with goodness of the fit, R^2 score 1 over a period of three months, which is a reasonable time horizon in the context of preventive maintenance. Furthermore, we validated the model and rectified the symptoms of the overfit by utilising the novel walk-forward algorithm; we yielded SoH predictions and the 95% point-wise confidence intervals for the predictions. Moreover, the Huang–Hilbert transformation of the data provides some means to analyze if the SoH starts to deteriorate, or, to change less frequently or more frequently than before, which may indicate point-of-time to change the model; the transformation needs to be performed dynamically as the length of the data series changes.

The future work could advance to extracting some user behavior patterns from the data as the number of the defined charging pulse timesteps (1000–2000) for each battery pack establish a basis for this kind of study. Also, an extension to this paper could be the further verification and development of the model for the battery SoH predictions with longer timeseries data, for example with the help of a relevant simulation, and further develop the model to make more robust predictions.

CRediT authorship contribution statement

Matti Huotari: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualisation, Supervision, Project administration, Funding acquisition. **Shashank Arora:** Conceptualisation, Verification, Funding acquisition. **Avleen Malhi:** Conceptualisation, Investigation, Writing - original draft, Writing - review & editing, Visualisation. **Kary Främling:** Conceptualisation, Resources, Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been supported by the European Commission through the H2020 project FINEST TWINS (grant no. 856602), by the Academy of Finland through the project grants 296096 and 322742, by European Commission through EUREKA Eurostars 2 (grant E! 12005 AIES) and by Business Finland. The authors appreciate the invaluable support of Professor Tanja Kallio and Mr. Panu Sainio from Aalto University in helping to evaluate the data and develop the research direction.

References

[1] K. Jalkanen, J. Karppinen, L. Skogström, T. Laurila, M. Nisula, K. Vuorilehto, Cycle aging of commercial NMC/graphite pouch cells at different temperatures, Appl. Energy 154 (2015) 160–172.
[2] L. Priyadarshini, P. Dash, S. Dhar, A new exponentially expanded robust random vector functional link network based MPPT model for local energy management of PV-battery energy storage integrated microgrid, Eng. Appl. Artif. Intell. 91 (2020) 103633.
[3] E. Martinez-Laserna, E. Sarasketa-Zabala, D.-I. Stroe, M. Swierczynski, A. Warnecke, J.-M. Timmermans, S. Goutam, P. Rodriguez, Evaluation of lithium-ion battery second life performance and degradation, in: 2016 IEEE Energy Conversion Congress and Exposition (ECCE), IEEE, 2016, pp. 1–7.
[4] C.F. Calvillo, A. Sánchez-Miralles, J. Villar, Energy management and planning in smart cities, Renew. Sustain. Energy Rev. 55 (2016) 273–287.
[5] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, K. Dietmayer, Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods, J. Power Sources 239 (2013) 680–688.
[6] S. Russell, P. Norvig, Artificial Intelligence: Pearson New International Edition: A Modern Approach, Pearson Higher Ed, 2013.
[7] G.E. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time Series Analysis: Forecasting and Control, Vol. 10, John Wiley & Sons, 2015, pp. 1–746.
[8] M. Huotari, Lithium-ion batteries timeseries data used in this work (to be published as the paper is published), https://github.com/huotarim/li-ion-battery-packs.
[9] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning the Elements of Statistical Learning, in: Springer Series in Statistics, vol. 27, New York, 2017, pp. 249–254.
[10] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. 30 (2017) 3146–3154.
[11] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42.
[12] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, Xgboost: extreme gradient boosting, 2015, pp. 1–4, R package version 0.4-2.
[13] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[14] B. Zhao, H. Lu, S. Chen, J. Liu, D. Wu, Convolutional neural networks for time series classification, J. Syst. Eng. Electron. 28 (1) (2017) 162–169.
[15] E. Schumann, Backtesting, in: M. Gilli, D. Maringer, E. Schumann (Eds.), Numerical Methods and Optimization in Finance, second ed., 2018.
[16] MATLAB, Rolling-window analysis of time-series models, 2021, URL https://se.mathworks.com/help/econ/rolling-window-estimation-of-state-space-models.html.
[17] U. Engineering, Building a backtesting service to measure model performance at uber-scale, 2021, URL https://eng.uber.com/backtesting-at-scale/.
[18] P.J. Luukko, J. Helske, E. Räsänen, Introducing libeemd: A program package for performing the ensemble empirical mode decomposition, Comput. Statist. 31 (2) (2016) 545–557.
[19] T.J. DiCiccio, B. Efron, et al., Bootstrap confidence intervals, Statist. Sci. 11 (3) (1996) 189–228.
[20] G. Hebrail, A. Berard, Individual Household Electric Power Consumption Data Set, EDF data at University of California, Irvine, School of Information and Computer Sciences, 2010, URL http://archive.ics.uci.edu/ml.
[21] M. Huotari, S. Arora, A. Malhi, K. Främling, A dynamic battery state-of-health forecasting model for electric trucks: Li-ion batteries case-study, in: ASME International Mechanical Engineering Congress and Exposition, Vol. 84560, American Society of Mechanical Engineers, 2020, V008T08A021.
[22] J.D. Kozlowski, Electrochemical cell prognostics using online impedance measurements and model-based data fusion techniques, in: 2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652), Vol. 7, IEEE, 2003, pp. 3257–3270.
[23] S. Makridakis, M. Hibon, The M3-competition: results, conclusions and implications, Int. J. Forecast. 16 (4) (2000) 451–476, http://dx.doi.org/10.1016/S0169-2070(00)00057-1.
[24] S. Makridakis, E. Spiliotis, V. Assimakopoulos, The M4 competition: Results, findings, conclusion and way forward, Int. J. Forecast. 34 (4) (2018) 802–808, http://dx.doi.org/10.1016/j.ijforecast.2018.06.001.
[25] M.A. Hannan, M.S.H. Lipu, A. Hussain, M.H. Saad, A. Ayob, Neural network approach for estimating state of charge of lithium-ion battery using backtracking search algorithm, IEEE Access 6 (2018) 10069–10079.
[26] D. Andre, A. Nuhic, T. Soczka-Guth, D.U. Sauer, Comparative study of a structured neural network and an extended Kalman filter for state of health determination of lithium-ion batteries in hybrid electricvehicles, Eng. Appl. Artif. Intell. 26 (3) (2013) 951–961.
[27] R.R. Richardson, M.A. Osborne, D.A. Howey, Gaussian process regression for forecasting battery state of health, J. Power Sources 357 (2017) 209–219.
[28] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, A.A. Lee, Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning, Nature Commun. 11 (1) (2020) 1–6.
[29] B. Saha, K. Goebel, Uncertainty management for diagnostics and prognostics of batteries using Bayesian techniques, in: 2008 IEEE Aerospace Conference, IEEE, 2008, pp. 1–8.
[30] C. Lei, J. Meng, D.-I. Stroe, J. Peng, G. Luo, R. Teodorescu, Multi-objective optimization of data-driven model for lithium-ion battery SOH estimation with short-term feature, IEEE Trans. Power Electron. (2020).
[31] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Remaining useful life estimation based on nonlinear feature reduction and support vector regression, Eng. Appl. Artif. Intell. 26 (7) (2013) 1751–1760.
[32] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, et al., Data-driven prediction of battery cycle life before capacity degradation, Nat. Energy 4 (5) (2019) 383–391.
[33] P.M. Attia, A. Grover, N. Jin, K.A. Severson, T.M. Markov, Y.-H. Liao, M.H. Chen, B. Cheong, N. Perkins, Z. Yang, et al., Closed-loop optimization of fast-charging protocols for batteries with machine learning, Nature 578 (7795) (2020) 397–402.
[34] S. Sepasi, R. Ghorbani, B.Y. Liaw, Improved extended Kalman filter for state of charge estimation of battery pack, J. Power Sources 255 (2014) 368–376.
[35] Battery pack design, validation,and assembly guide using A123 systems AMP20M1HD-A nanophosphate® cells, 2014, URL https://formula-hybrid.org/wp-content/uploads/A123_AMP20_battery_Design_guide.pdf.
[36] Y. Zhang, Y. Li, Y. Tao, J. Ye, A. Pan, X. Li, Q. Liao, Z. Wang, Performance assessment of retired EV battery modules for echelon use, Energy 193 (2020) 116555, http://dx.doi.org/10.1016/j.energy.2019.116555.
[37] Y. Hua, A. Cordoba-Arenas, N. Warner, G. Rizzoni, A multi time-scale state-of-charge and state-of-health estimation framework using nonlinear predictive filter for lithium-ion battery pack with passive balance control, J. Power Sources 280 (2015) 293–312.
[38] H. Ren, Y. Zhao, S. Chen, T. Wang, Design and implementation of a battery management system with active charge balance based on the SOC and SOH online estimation, Energy 166 (2019) 908–917.
[39] Y.C. Zhang, O. Briat, J.-Y. Delétage, C. Martin, N. Chadourne, J.-M. Vinassa, Efficient state of health estimation of li-ion battery under several ageing types for aeronautic applications, Microelectron. Reliab. 88–90 (2018) 1231–1235, 29th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF 2018).
[40] L. Wang, C. Pan, L. Liu, Y. Cheng, X. Zhao, On-board state of health estimation of LiFePO4 battery pack through differential voltage analysis, Appl. Energy 168 (2016) 465–472.
[41] B. Saha, K. Goebel, Battery data set, in: NASA AMES Prognostics Data Repository, NASA Ames, 2007.
[42] I. Pratama, A. Permanasari, I. Ardiyanto, R. Indrayani, A review of missing values handling methods on time-series data, 2016, pp. 1–6, http://dx.doi.org/10.1109/ICITSI.2016.7858189.
[43] Z. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu, Recurrent neural networks for multivariate time series with missing values, Sci. Rep. 8 (1) (2018) 1–12.
[44] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 454 (1971) (1998) 903–995.
[45] N.R. Babu, B.J. Mohan, Fault classification in power systems using EMD and SVM, Ain Shams Eng. J. 8 (2) (2017) 103–111.
[46] Z. Wu, N.E. Huang, S.R. Long, C.-K. Peng, On the trend, detrending, and variability of nonlinear and nonstationary time series, Proc. Natl. Acad. Sci. 104 (38) (2007) 14889–14894.
[47] Z. Yang, C. Bingham, B.W.-K. Ling, M. Gallimore, P. Stewart, Y. Zhang, Trend extraction based on Hilbert-Huang transform, in: 2012 8th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP), 2012, pp. 1–5, http://dx.doi.org/10.1109/CSNDSP.2012.6292713.
[48] J.C. Rawlins, CHAPTER 13 - Complex RLC circuit analysis, in: J.C. Rawlins (Ed.), Basic AC Circuits (Second Edition), second ed., Newnes, Burlington, 2000, pp. 427–452.

[49] A. Baccigalupi, A. Liccardo, The huang Hilbert transform for evaluating the instantaneous frequency evolution of transient signals in non-linear systems, Measurement 86 (2016) 1–13.
[50] B. Boashash, Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals, Proc. IEEE 80 (4) (1992) 520–538.
[51] R. Yan, R.X. Gao, Hilbert–Huang transform-based vibration signal analysis for machine health monitoring, IEEE Trans. Instrum. Meas. 55 (6) (2006) 2320–2329.
[52] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 785–794, http://dx.doi.org/10.1145/2939672.2939785.
[53] W. Yucong, W. Bo, Research on EA-xgboost hybrid model for building energy prediction, in: Journal of Physics: Conference Series, Vol. 1518, IOP Publishing, 2020, 012082.
[54] A. Dineva, B. Csomós, S. Kocsis Sz., I. Vajda, Investigation of the performance of direct forecasting strategy using machine learning in state-of-charge prediction of li-ion batteries exposed to dynamic loads, J. Energy Storage 36 (2021) 102351.
[55] D. Zwillinger, S. Kokoska, CRC Standard Probability and Statistics Tables and Formulae, Crc Press, 1999.
[56] C. Sammut, G.I. Webb (Eds.), Mean squared error, in: Encyclopedia of Machine Learning, Springer US, Boston, MA, 2010, p. 653.
[57] C. Sammut, G.I. Webb (Eds.), Mean absolute error, in: Encyclopedia of Machine Learning, Springer US, Boston, MA, 2010, p. 652.
[58] Coefficient of determination, in: The Concise Encyclopedia of Statistics, Springer New York, New York, NY, 2008, pp. 88–91.
[59] T. pandas development team, Pandas-dev/pandas: Pandas, 2020, http://dx.doi.org/10.5281/zenodo.3509134.
[60] S. Simpson, Foundation of Mathematics, The Pennsylvania State University, University Park, Pennsylvania, 2009.
[61] E. Zivot, J. Wang, Vector autoregressive models for multivariate time series, in: Modeling Financial Time Series with S-Plus®, Springer, 2006, pp. 313–359.
[62] M. Fleischhammer, T. Waldmann, G. Bisle, B.-I. Hogg, M. Wohlfahrt-Mehrens, Interaction of cyclic ageing at high-rate and low temperatures and safety in lithium-ion batteries, J. Power Sources 274 (2015) 432–439, http://dx.doi.org/10.1016/j.jpowsour.2014.08.135.
[63] S. Arora, Selection of thermal management system for modular battery packs of electric vehicles: A review of existing and emerging technologies, J. Power Sources 400 (2018) 621–640, http://dx.doi.org/10.1016/j.jpowsour.2018.08.020.
[64] S. Arora, A. Kapoor, W. Shen, A novel thermal management system for improving discharge/charge performance of Li-ion battery packs under abuse, J. Power Sources 378 (2018) 759–775.
[65] D. Rey, M. Neuhäuser, Wilcoxon-signed-rank test, in: International Encyclopedia of Statistical Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1658–1659.
[66] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and machine learning forecasting methods: Concerns and ways forward, PLoS One 13 (3) (2018) e0194889.
[67] S.C. Harris, Hybrid vehicle with modular battery system, 2017, US Patent 9, 579, 961.
[68] A. Hefnawy, T. Elhariri, C. Cherifi, J. Robert, A. Bouras, S. Kubler, K. Främling, Combined use of lifecycle management and IoT in smart cities, in: 2017 11th International Conference on Software, Knowledge, Information Management and Applications (Skima), IEEE, 2017, pp. 1–9.