A novel semi-supervised learning approach for maritime lithium-ion battery monitoring

Clara Bertinelli Salucci[1], Azzeddine Bakdi[2], Ingrid Kristine Glad[1], Erik Vanem[13], Riccardo De Bin[1]

E-mail for correspondence: clarabe@math.uio.no

Abstract: We propose a novel semi-supervised learning method to monitor the State of Health of lithium-ion batteries, a prominent technology for the electrification of the transport sector. Our approach enables State of Health monitoring of batteries with no labeled data, starting from a minimal set of labeled data from another similar battery. This can be achieved by exploiting the relation between a pseudo-capacity measure and the total capacity of the labeled data. Our results with operational data from maritime batteries show that the approach is valid and can lead to significant progress in failure prevention, operational optimization, and for planning batteries at the design stage.

Keywords: Semi-supervised learning; Multivariable Fractional Polynomials; Li-ion battery State of Health.

1 Motivating Problem and Data Description

Monitoring of the State of Health (SoH) of lithium-ion (Li-ion) batteries is crucial for maritime applications. In fact, over time and over usage Li-ion batteries undergo ageing mechanisms that ultimately lead to battery failure; the consequences for a vessel at sea can be potentially catastrophic, therefore it is compelling to assess the battery conditions with good accuracy.

One way of quantifying the SoH is based on the degradation of the battery capacity (Vanem et al., 2021):

$$\mathrm{SoH}_i = \frac{C_{\text{available}}}{C_{\text{nominal}}} \times 100\,(\%), \tag{1}$$

where $C_{\text{available}}$ is the actually available capacity, and C_{nominal} is the nominal capacity of the battery. However, estimating the capacity itself is often a

demanding exercise. It is common practice in the maritime field to carry out annual tests, that enable an estimate of the battery capacity. Such tests, however, are burdensome and time-consuming, and provide very sparse capacity assessments. Further, as they are performed under different conditions (temperature, durations, etc.), they can hardly be related to each other, resulting to some extent inaccurate. An efficient method for continuous battery diagnostic, thus, is strongly needed. Data-driven methods can be greatly advantageous as they are agnostic to the real and highly complex physical problem, they are ductile and can be used for different batteries (Vanem et al., 2021).
Operating data for this analysis are provided by a leading supplier of energy storage systems for maritime applications. The data pertain to the battery systems of three vessels, which we regard as three different datasets. For each of them, we have high-frequency sensor data: temperature, voltage and current intensity measurements, together with the battery State of Charge (SoC) which we regard as sensor data insofar as it is provided by the company with good accuracy. Such variables are continuously measured from the beginning of operation until 4.5 or 5.5 years later, though there are periods of missing data in all datasets.
Our minimal set of labeled data consists of three data-points for one of the vessels, obtained from three SoH tests conducted in years that are not necessarily consecutive. Our approach is based on relating the discharge phases of the batteries while they age over years: thus, we pre-process the data to go from continuous measurements series to single events, the discharge cycles, identified on the basis of changes of sign in the SoC derivative.

2 Semi-Supervised Learning Methodology

We will refer to the three datasets according to the following:

- Reference data: data from the dataset with the three labeled data-points (vessel A);
- Target data: data from the other two datasets (vessels B and C).

Consequently, the labeled cycles from vessel A will be called *reference cycles*, while all cycles from the other two vessels are *target cycles*; our aim is to predict the total capacity of the battery at the target cycles.
Our approach relies on a fundamental assumption: the SoH can be considered constant for a time window around the day where the measurement was taken. The assumption is valid as the SoH is known to degrade gently and almost linearly in the first years of operations, after an initial short stage in which the degradation is more pronunced, and before a final stage where the decay is faster and non-linear (Edge et al., 2021). This enables us to enlarge the set of reference cycles. The method develops in three steps:

1. Cycle classification: using a tree-like classification, cycles with similar characteristics are grouped together, and each class is treated independently. This is important to account for the large impact that different conditions (temperature, SoC range, C-rate, etc.) have on the estimated capacity. At the end of this step, the data are organised in tables containing cycles with similar characteristics. An example is provided in Table 1, where the first two cycles are from the reference dataset, and hence they have an estimated SoH, and three other similar cycles in datasets B and C have been matched.

2. Model training: using the reference cycles, we train a linear model in each class. The total capacity of the battery as from the SoH test is our dependent variable; the number of features entering the model depends on how many reference cycles we have in the considered class. In all cases we input the *pseudo-capacity*,

$$\tilde{C} = \int_{t_{\text{start}}}^{t_{\text{end}}} I(t)\,\mathrm{d}t; \tag{2}$$

   optionally, cycle characteristics such as duration, initial time, variance in the C-rate and temperature etc. are also included in the model. We discard all classes having models with $R^2 < 0.6$.

3. Total capacity estimation: in each class, we get capacity estimates for all target cycles from the model trained at step 2.

The capacity estimates from different classes are then gathered together and converted to the SoH scale. In real applications it is often convenient to have weekly or monthly SoH estimates, therefore we do a weighted average of the estimations where the weights are the reciprocal of the uncertainties estimated by the model. This is done in order to ensure that highly uncertain estimates contribute very little to the final estimate.

TABLE 1. Example of a few cycles from the same class: the first two rows are cycles from the reference dataset, and hence they have an estimated SoH. Other three similar cycles in datasets B and C have been matched and are a target for capacity estimation.

SoC_1	SoC_2	avg_cRate	max_temp	min_temp	SoH	dataset
89%	**73%**	**-0.372**	**28°**	**24°**	**92.4%**	**A**
89%	**73%**	**-0.379**	**27°**	**23°**	**92.4%**	**A**
90%	73%	-0.374	27°	24°	–	B
89%	72%	-0.379	27°	23°	–	B
89%	71%	-0.377	27°	24°	–	C

3 Multivariable Fractional Polynomials for SoH modelling

The semi-supervised approach provides SoH estimates which transform the large unlabelled datasets into training data for modelling the battery degradation: the Multivariable Fractional Polynomials (MFP) approach (Sauerbrei and Royston, 1999) has been chosen for the purpose, in view of the encouraging results achieved on lab data in a previous work (Bertinelli Salucci et al., 2022). The response variable of the model is the monthly change in the battery SoH with respect to the initial value $\mathrm{SoH}(t_0)$,

$$y = \Delta\mathrm{SoH}(t) = \mathrm{SoH}(t_0) - \mathrm{SoH}(t), \tag{3}$$

while the set of candidate covariates is derived from the battery sensor data for all charge and discharge cycles, including a few significative interaction terms. All features are cumulative over each month (e.g. sum of durations of charge or discharge phases, average C-rates, ...), except for the *equivalent full cycles* measure (efc) which is cumulative over the whole history of the battery system. The MFP algorithm selects the most suitable polynomial transformations of the covariates among a set of possible choices, and variable selection is also performed (significance level $P < 0.05$) to achieve potential variance reduction and ease the model interpretability. The regression model has been trained on data from vessel B and tested on vessel C.

4 Results and Conclusions

Monthly averaged results obtained with the semi-supervised learning approach are shown in Figure 1 and Figure 2 for the two target ships. The unavailability of frequent and reliable labels makes it difficult to provide a specific accuracy assessment for the method; however, our results are in line with the typical degradation patterns of Li-ion batteries depicted by Edge et al. (2021), as well as with battery experts' expectations.
The left panel of Figure 3 shows the results obtained in predicting the SoH degradation of vessel C with the MFP model trained on data from vessel B (Table refbertinellisalucci:tab2). The plot confirms the effectiveness of MFP regression for modelling SoH degradation of lithium-ion batteries: the predicted values are all very close to the estimates obtained with the semi-supervised approach, with a normalized Root Mean Squared error of 0.85%. The right panel of the figure presents an histogram of the normalised absolute error: most of the errors are below 1.5%, and all errors below 2%.

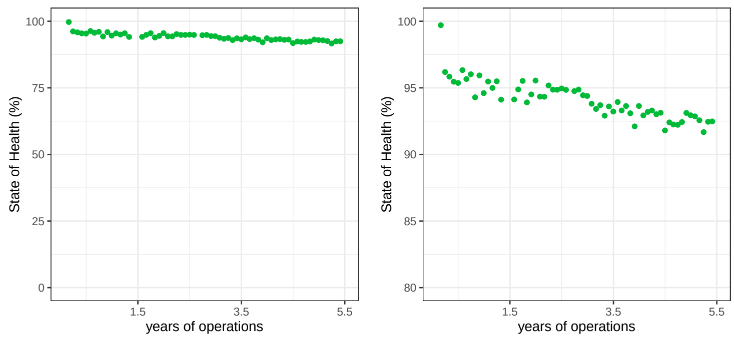

FIGURE 1. Montly averaged SoH estimates for vessel B on full scale (left) and reduced scale 80-100% (right).

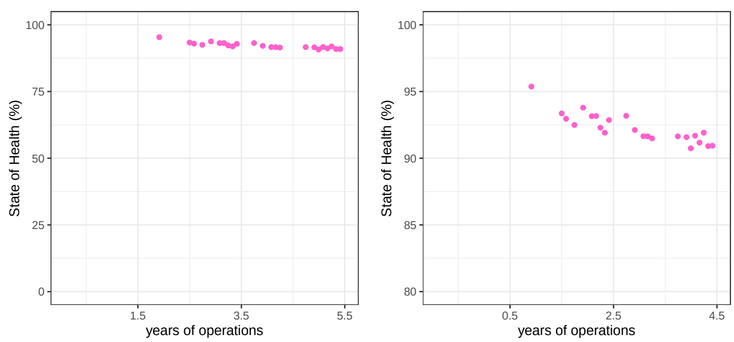

FIGURE 2. Montly averaged SoH estimates for vessel C on full scale (left) and reduced scale 80-100% (right).

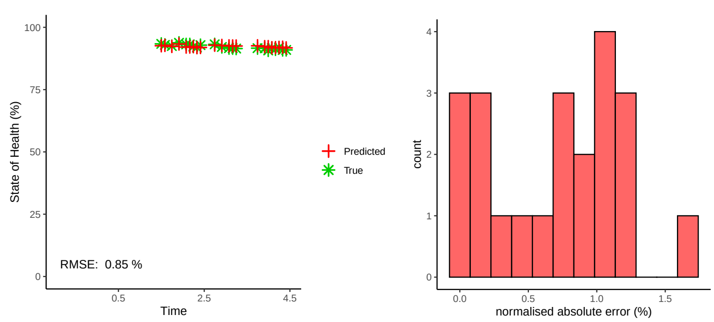

FIGURE 3. Left: State of Health degradation results for vessel C using the MFP regression model. Right: histogram of the normalised absolute error.

TABLE 2. MFP regression model trained on data from vessel B. The features entering the model after the variable selection mechanism are reported together with their estimated coefficients, standard errors and corresponding p-values: efc is a measure of the equivalent full cycles of the battery; $V_{\text{in,disch.}}$ is the average initial voltage of the discharges cycles in one month; $T_{\text{min},3}$ and $T_{\text{min},1}$ are the monthly averages of the minimum values of two temperature sensors.

	est. coefficient	**std. error**	***p*-value**
Intercept	-6.395	1.62	0.0002
$efc/10^5$	83.31	20.91	0.0002
$V_{\text{in,disch.}} : efc/10^5$	-16.53	4.42	0.0005
$T_{\text{min},3}/10$	-26.19	8.25	0.0025
$T_{\text{min},1}/10$	30.14	8.93	0.0014

References

Bertinelli Salucci, C., Bakdi, A., Glad, I.K., Vanem, E., and De Bin, R. (2022). Multivariable Fractional Polynomials for lithium-ion batteries degradation models under dynamic conditions. *Journal of Energy Storage*, **52**, 104903.

Edge, J., O'Kane, S., Prosser, R., Kirkaldy, et al. (2021). Lithium Ion Battery Degradation: What you need to know. *Physical Chemistry Chemical Physics*, **23**.

Sauerbrei W. and Royston P. (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **162**, 71-94.

Vanem, E., Bertinelli Salucci, C., Bakdi, A., and Alnes, Ø. Å. (2021). Data-driven state of health modelling—A review of state of the art and reflections on applications for maritime battery systems. *Journal of Energy Storage*, **43**, 103158.