Data-Driven Approaches to Diagnostics and State of Health Monitoring of Maritime Battery Systems

Erik Vanem[1, 2], Qin Liang[1], Carla Ferreira[1], Christian Agrell[1], Nikita Karandikar[1], Shuai Wang[1], Maximilian Bruch [3], Clara Bertinelli Salucci[2], Christian Grindheim[2], Anna Kejvalova[2], Øystein Åsheim Alnes [4], Kristian Thorbjørnsen[5], Azzeddine Bakdi[5], Rambabu Kandepu[5]

[1] *DNV Group Research & Development, Høvik, Norway*
Erik.Vanem@dnv.com
Qin.Liang@dnv.com
Carla.Ferreira@dnv.com
Christian.Agrell@dnv.com
Nikita.Karandikar@dnv.com
Shuai.Wang@dnv.com

[2] *Department of Mathematics, University of Oslo, Oslo, Norway*
erikvan@math.uio.no
clarabe@math.uio.no
chrgrindheim@gmail.com
annakej@student.ikos.uio.no

[3] *Fraunhofer ISE, Freiburg, Germany*
maximilian.bruch@ise.fraunhofer.de

[4] *DNV Maritime, Høvik, Norway*
oystein.alnes@dnv.com

[5] *Corvus Energy, Porsgrunn, Norway*
kthorbjornsen@corvusenergy.com
abakdi@corvusenergy.com
rkandepu@corvusenergy.com

Abstract

Battery systems are increasingly being used for powering ocean going ships, and the number of fully electric or hybrid ships relying on battery power for propulsion and maneuvering is growing. In order to ensure the safety of such electric ships, it is important to monitor the available energy that can be stored in the batteries, and classification societies typically require that the state of health (SOH) can be verified by independent tests. However, this paper addresses data-driven approaches to state of health monitoring of maritime battery systems based on operational sensor data. Results from various approaches to sensor-based, data-driven degradation monitoring of maritime battery systems will be presented, and advantages and challenges with the different methods will be discussed. The different approaches include cumulative degradation models and snapshot models. Some of the models need to be trained, whereas others need no prior training. Moreover, some of the methods only rely on measured data, such as current, voltage and temperature, whereas others rely on derived quantities such as state of charge (SOC). Models include simple statistical models and more complicated machine learning techniques. Different datasets have been used in order to explore the various methods, including public datasets, data from laboratory tests and operational data from ships in actual operation. Lessons learned from this exploration will be important in establishing a framework for data-driven diagnostics and prognostics of maritime battery systems within the scope of classification societies.

1. INTRODUCTION

The safety of battery-powered ships is important, and this is verified and ensured by classification services. Fire and explosion are obvious risks, but another central aspect is ensuring that the available energy stored in the batteries is sufficient to cover the required propulsion or manoeuvring power demand (Hill et al., 2015). Loss of propulsion power in a critical situation can lead to serious accidents such as collision or grounding. Therefore, reliable estimation and prediction of the actual available energy of a battery is crucial for ship safety.

Battery systems are aging, meaning that the energy storage capacity degrades by calendar time and by charge/discharge cycles (Pop et al., 2008). The aging process affects both the amount of charge that can be stored and the available power. Most maritime battery systems are designed with an expected lifetime of 10 years and end of life is typically defined as State of Health (SOH) = 70-80%, where SOH stands for the ratio of remaining capacity to initial capacity (in %). Ships, on the other hand, are typically design for 25-30 years. Hence, batteries are expected to approach their end of useful life (EOL) long before the end of the operational life of the vessels. In such a context, reliable estimation of SOH will become increasingly important as the battery systems approaches its EOL and making correct decisions on remaining useful life (RUL) will have great financial and safety implications.

Currently within the maritime Industry, the governing methodology for evaluating the real-time capacity of a battery system on board a fully electric or battery-hybrid vessel is by considering the State of Charge (SOC) and the State of Health (SOH) in the following simplistic way: Available energy = Initial Capacity x SOC x SOH. A major part of such an estimation will be a reliable evaluation of the battery State of Health. Currently, battery suppliers are required to have an SOH estimation algorithm and to verify the SOH annually through in-situ capacity testing. From a practical point of view, the annual capacity test is time consuming and typically requires that the ship is taken out of operation for one full day. Moreover, the accuracy of the test is questionable due to several factors influencing the results, such as variability in loads, temperatures and Depth of Discharge (DOD). As ship-to-shore connectivity has improved immensely over the past few years it is natural to evaluate whether a sensor-based monitoring system can both reduce downtime for the operator and improve the quality of the SOH verification.

This paper summarizes and presents several approaches to data-driven modeling of state of health of maritime battery systems. The overall idea is to use sensor data from batteries to learn the degradation state of the batteries without the need for specific testing or characterization cycles. If successful, data-driven approaches may replace the need for annual capacity testing to verify SOH according to class rules.

A review of different approaches for data-driven diagnostics of maritime battery systems were presented in Vanem, Bertinelli Salucci, et al. (2021); Vanem, Alnes, & Lam (2021). According to this review, data-driven methods for estimating battery capacity can be categorized into a few generic type of approaches. Additionally, a distinction was made between models that rely on the complete loading history of the batteries in order to estimate current state of health and what was referred to as snapshot methods, where state of health and capacity can be estimated based on only snapshots of the data. In this paper, the exploration of different approaches and methods will be outlined, including both snapshot methods and cumulative methods.

It is important to note that different approaches set different requirements for the data. For example, most approaches require training data to train the data-driven models, whereas some approaches can do without training data. If training data are required, they obviously need to be representative of typical operational data, and should preferably correspond to identical cells as the system it should be applied to. This is among the considerations that need to be made when comparing and recommending which models to use for data-driven classification of the batteries. Other factors to consider, apart from the predictive performance, include amount of data needed, the sensitivity to missing data and the computational costs.

2. BATTERY DATASETS

Different sets of data have been available for analysis and modeling in the project. These include laboratory data generated by the project, proprietary data from actual ships in operation and some publicly available data.

2.1. Laboratory Test Data

Three different types of battery cells have been subject to cycling tests at Fraunhofer's laboratory in order to generate degradation data for the cells. Two types of cylindrical 18650 cells, i.e. energy cells (henceforth denoted DDE; nominal capacity 3.5 Ah) and power cells (henceforth denoted DDP; nominal capacity 2.5 Ah), and one type of pouch cells (henceforth denoted DDF; nominal capacity 64 Ah) have been cycled according to specified test matrices. Individual cells have been cycled within specified lower and upper voltage limits, with specified charge and discharge currents, and at specified controlled temperatures. Varying these parameters yields different degradation rates. This continuous cycling is interrupted at regular intervals to perform check-ups and capacity measurements, i.e. pulse tests and charge and discharge capacity measurements by way of Coulomb counting over deep cycles at low current rates. Hence, capacities will be measured at certain points in time for all cells. Results are illustrated in Figure 1, which show estimated capacity as a

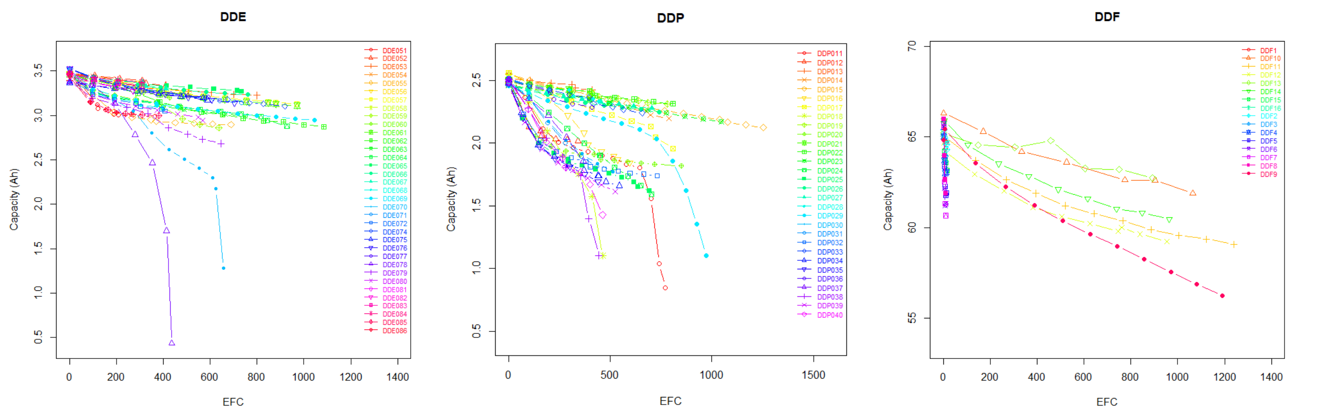

Figure 1. Capacity versus equivalent full cycles from Fraunhofer laboratory tests; DDE (left), DDP (middle) and DDF cells

function of the number of equivalent full cycles (EFC) for the three types of cells. EFC has been calculated based on current integration compared to the nominal capacity. It is noted that some of the DDF cells have not been cycled - they are simply stored over time. Hence, a significant drop in capacity is observed without any increase in EFC for some of the DDF cells, and this is due to calendar aging (rightmost plot in Figure 1).

Values of current, voltage and temperature are sampled continuously, resulting in time-series of these variables throughout the experiment. From these raw measurements, different derived variables can be calculated as well, such as cumulative throughputs, cycle counts and equivalent full cycles. Measurements are obtained from a total of 81 individual cells; 35 DDE cells, 30 DDP cells and 16 DDF cells. The cells in this experiments have been charged and discharged according to a constant-current-constant voltage (CCCV) scheme: the cells are charged/discharged with constant current until the cut-off voltage, where the cells continue to charge/discharge at constant voltage with a current that gradually decreases towards zero.

A similar set of data from cycling tests performed on the same cell types but at a different lab at Corvus' has been made available. Two types of cylindrical cells have been tested, corresponding to energy cells and power cells, which are similar to the DDE and DDP cells tested at Fraunhofer. Again, the cells have been subject to repeated cycling, with regular capacity measurements by Coulomb counting. Figure 2 shows the measured remaining capacity as a function of equivalent full cycles (EFC) for the energy (E) and power (P) cells, respectively. The figures indicate both charge and discharge capacities, measured during charging and discharging tests, and these are very similar, with only slightly higher values for the charge capacity. The regular cycling in these tests are slightly different from the tests performed at Fraunhofer. That is, the regular cycling consists of repeatedly charging the cells with a constant current - constant voltage scheme (CCCV), then discharging with a constant power, with small pauses in between.

Some additional data have been available from DNV's lab testing facilities, on battery cells of similar types as the ones used in the operational data. For some applications, these have been used as training data to train the data-driven models.

2.2. Operational Data From Ships in Service

Field data from electric ships with a battery system of pouch cells of type DDF have been analyzed in this study. These battery systems are designed with a 4-layer structure; individual *cell-pairs* connected in series make up *modules*, modules connected in series form *packs* and several packs connected in parallel make up an *array*. A ship may have one or more arrays connected in parallel as independent energy storage systems that do not communicate directly, and any combination of packs in an array can be powered off during operation. Raw data from these systems include the pack voltage and current for all packs as well as the cell voltage, temperature and State of Charge (SOC) for all cells (or rather, cell-pairs). Since modules and cell-pairs are connected in series within a pack, the current will be the same for all series elements in that pack. However, it will not be possible to distribute this current over the two cells in a series element. Hence, for all practical purposes, the cell-pairs will be considered the smallest entities of the system, i.e. cells. Moreover, whereas currents, voltages and temperatures are measured directly by sensors, SOC is a derived quantity that needs to be calculated from the other raw sensor measurements. An example of time-series from this system is shown in Figure 3.

Operational data from 6 different ships with the same battery system on board have been available for this study. These ships include both hybrid and all-electric solutions and with different configurations (different number of arrays and packs

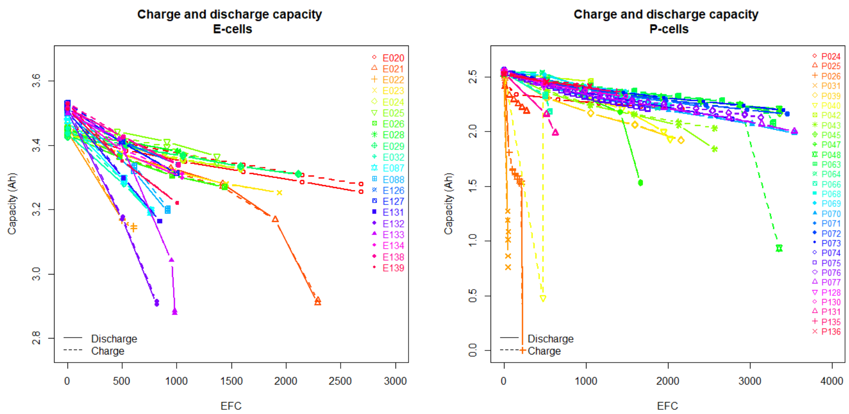

Figure 2. Capacity as a function of equivalent full cycles from Corvus laboratory tests; energy (left) and power (right) cells

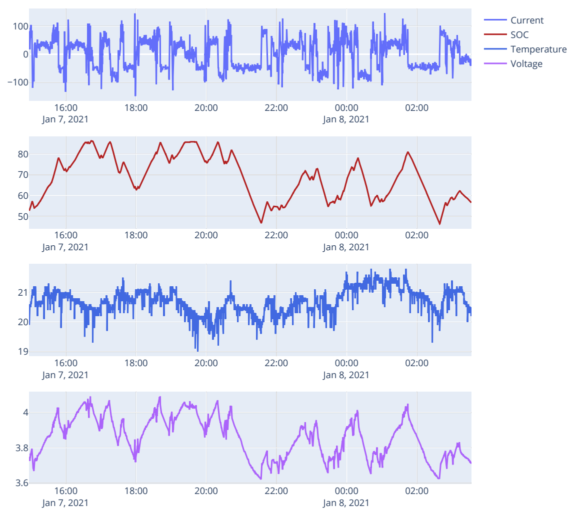

Figure 3. Example of time-series from the onboard battery system; pack current (top), cell SOC (2nd row), cell temperature (3rd row) and cell voltage (bottom)

per array). All of these systems are relatively new, without having experienced extensive degradation. Results from annual capacity tests are available, and all vessels had undergone at least two such tests by the time of this study. The ships battery system configuration can be found in Table 1. Trace plots of the operational data from one cell on Vessel C is shown in Figure 4. It is observed that there is a rather long gap in the data, and also that this system started doing fast-charging at some point, with a sudden occurrence of higher maximum currents.

Additional data from an older battery system have been analyzed in this study. These have different types of battery cells, but with the benefit of longer time histories. However, data quality is somewhat lower, and there are more and longer data gaps compared to the newer system. Data from these systems

Table 1. Ships battery system configuration

Ship ID	Battery system	# arrays	# packs per array	# Annual tests
Ship A	Hybrid	2	4	2
Ship B	Hybrid	1	12	3
Ship C	All-electric	2	9	2
Ship D	All-electric	2	9	2
Ship E	All-electric	2	9	2
Ship F	Hybrid	2	13	3

are referred to as Site A, Site B and Site C and include data from three different vessels.

2.3. Publicly Available Datasets

There are a number of publicly available datasets from various types of batteries (dos Reis et al. (2021)). Some of these contain degradation data from cycling tests, and has been utilized in this study.

Several battery datasets have been made publicly available by the NASA Ames Prognostics Center of Excellence (PCoE), of which a randomized battery usage dataset has be analyzed in this study Bole et al. (2014)[1]. It consist of aging data of 18650 lithium-ion batteries under randomly generated usage profiles. The loading of these batteries contain two regimes: reference discharges and random walk steps. The reference discharge step is a controlled full discharge cycle following a controlled full charge cycle. This is used to estimate the cell capacity periodically. Both the charge and the discharge in these cycles follow the CCCV principle. During the random walk steps, a charge or discharge current is randomly sampled from a set of current values, and then kept constant for a predetermined period of time or until a voltage limit is reached. Then, a new current is drawn and this continues until a new

[1]Bole, B. and Kulkarni, C. and Daigle, M. *Randomized Battery Usage Data Set.* NASA Arnes Prognostics Data Repository. URL: http://ti.arc.nasa.gov/project/prognostic-data-repository

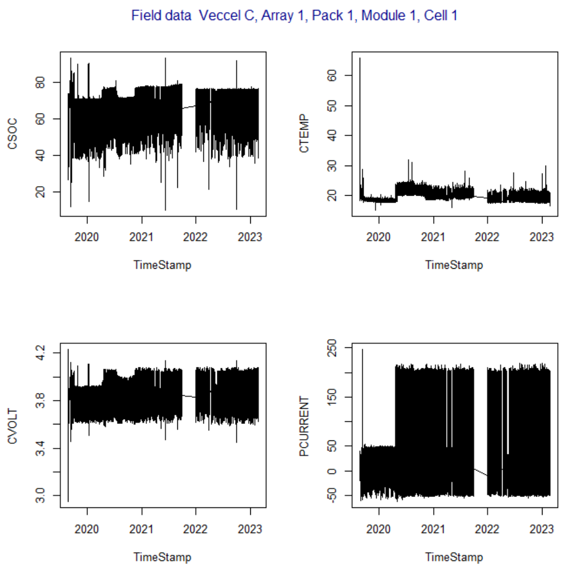

Figure 4. Raw data from one cell from Vessel C

reference cycle is performed at pre-determined intervals. At all times, values of temperature, current and voltage are measured. For more details on these data, reference is made to Bole et al. (2014), see also Bertinelli Salucci et al. (2022).

3. DATA-DRIVEN STATE OF HEALTH ESTIMATION

In this section of the paper, a description of various data-driven approaches that have been explored in this study will be presented. They include both cumulative damage models, snapshot methods and other approaches, and are applied to different datasets.

3.1. Cumulative Degradation Methods

3.1.1. Battery.ai

Battery AI is an artificial intelligence driven, semi-empirical battery analytics platform (Xue et al. (2022))[2]. It provides real-time battery health analysis considering both cycling and calendar aging. The aim of this case study is to test the capability, functionality and prediction accuracy of Battery AI on real operational data. Battery AI uses a combination of machine learning and semi-empirical methods to model battery behavior under various real-world conditions. The platform can analyze complex duty cycles in real-world operating conditions and determining the constituent abuse factors. The impact of these abuse factors is then modeled to predict total degradation. In short, a semi-empirical function on the form

$$y = 100 - A \times TO^p + C \tag{1}$$

is fitted to degradation data, where TO denotes turnover and C is a calibration factor. The exponent p is estimated from empirical degradation curves, and A models the contribution from the combined effect of the various stress factors by way of a neural network trained on cycling test data. Detailed description of this tool is out of scope of this paper, but references is made to Xue et al. (2022); see also an extended description of this case study in Liang et al. (2022, 2023). Figure 5 shows an example of results from this method on one cell from one of the vessels investigated in this study. Results are reasonably good, with deviations from the annual capacity tests in the order of 3% after one and a half year. Note also that estimated SOH > 100% means that the estimated capacity is greater than the nominal capacity. Results for the other vessels are similar, but it is stressed that verification is difficult without longer time series of data and more degradation.

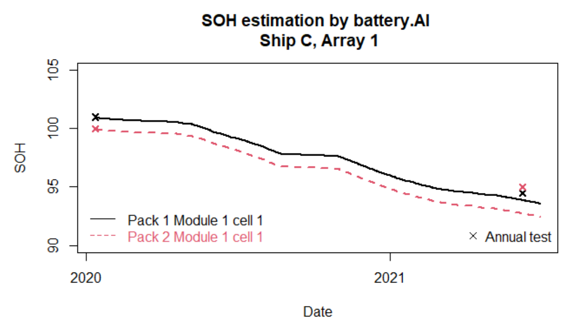

Figure 5. Example of SOH predictions from battery.AI compared to annual test

One critical step in analyses using the battery.AI tool is the pre-processing of the data to get it on the required input format. This includes cycle decomposition of the raw data to identify cycles and assign stress factor to these. These are then combined in a cumulative way, where degradation contributions from the individual cycles are added together. As elaborated in Liang et al. (2023), this is both time-consuming and expensive when the amount of data grows large. For large maritime battery systems, which contains several thousand cells, and are operated over several years, it was found that the amount of data handling and pre-processing is a significant barrier to practical large-scale implementation of this tool. Moreover, since it is a cumulative method, it relies on the complete data stream and do not handle gaps in the data very well. Nevertheless, results from this case study indicate that the battery.AI tool might be able to predict battery degradation reasonably well. Notwithstanding, the practical limitations with regards to the data handling and the fact that it is a cumulative method that requires complete time series is a barrier for widespread implementation on a fleet of ships.

[2]Battery.AI; url: https://www.dnv.com/services/battery-ai-35181
Battery AI on Veracity; url: https://store.veracity.com/battery-ai

3.1.2. Probabilistic cumulative models

This section describes a way to model the battery state of health (SOH) for varying loading conditions, using data from controlled laboratory experiments as training data and applied to predict SOH on operational data from ships. In the controlled laboratory experiments, cells are cycled under fixed loading conditions, and the SOH as a function of the number of cycles is estimated. A loading condition is represented by the following five parameters: 1) Charge C-rate, 2) Discharge C-rate, 3) Upper state of charge, 4) Lower state of charge, and 5) Temperature.

We consider inputs of the form $(\mathbf{x}, n)$ where $\mathbf{x} \in \mathbb{R}^5$ and $n \in \mathbb{N}$. Here $\mathbf{x} = (x_1, \ldots, x_5)$ represents a loading scenario ('C-rate charge', 'C-rate discharge', 'SOC upper', 'SOC lower', 'Temp'), and n is the number of cycles. Hence, $(\mathbf{x}, n)$ means that n cycles are performed under condition $\mathbf{x}$. Assume that the cell is subjected to n_1 cycles under condition $\mathbf{x}_1$, followed by n_2 cycles under condition $\mathbf{x}_2$, and so on. We will write this as $(\mathbf{X}, \mathbf{n})$ where $\mathbf{X} = (\mathbf{x}_1, \ldots \mathbf{x}_k)$ and $\mathbf{n} = (n_1, \ldots, n_k)$. Hence, $(\mathbf{X}, \mathbf{n})$ represents a total of $\sum_{i=1}^{k} n_i$ cycles under the k different conditions $\mathbf{x}_1, \ldots, \mathbf{x}_k$.

Let $f_{\mathbf{x}}(n)$ denote the SOH after n cycles under condition $\mathbf{x}$. We assume that all cells start with 100% SOH, i.e. $f_{\mathbf{x}}(0) = 1$ for all $\mathbf{x}$. Let $S(\mathbf{X}, \mathbf{n})$ denote the SOH after loading $(\mathbf{X}, \mathbf{n})$. That is, the SOH after the cell has been subjected to n_1 cycles under condition $\mathbf{x}_1$, followed by n_2 cycles under condition $\mathbf{x}_2$, and so on. We have that $S(\cdot, \mathbf{0}) \equiv 1$. We will also define the *reduction in SOH* which we denote Δf. Let $\Delta f(s, \mathbf{x}, n)$ be the reduction in SOH after starting at $SOH = s$ and then run n cycles under condition $\mathbf{x}$. When the condition $\mathbf{x}$ is fixed, we have that

$$\Delta f(s, \mathbf{x}, n) = s - f_{\mathbf{x}}\left(f_{\mathbf{x}}^{-1}(s) + n\right). \tag{2}$$

To model the SOH in this way, we rely on a "Markovian" assumption of path independence, i.e. we assume that the total reduction in SOH corresponding to the variable loading $(\mathbf{X}, \mathbf{n})$ is the sum of the SOH reductions corresponding to the individual scenarios $(\mathbf{x}_1, n_1), \ldots, (\mathbf{x}_k, n_k)$. That is,

$$S(\mathbf{X}, \mathbf{n}) = 1 - \sum_{i=1}^{k} \Delta s_i, \tag{3}$$

where

$$\begin{aligned} \Delta s_i &= \Delta f(s_{i-1}, \mathbf{x}_i, n_i), \quad s_0 = 1, \\ s_{i-1} &= 1 - (\Delta s_1 + \cdots + \Delta s_{i-1}). \end{aligned} \tag{4}$$

From N controlled experiments we obtain a dataset $D = \{(\mathbf{x}_i, \mathbf{n}_i, \mathbf{s}_i)\}_{i=1}^{N}$. Here $\mathbf{x}_i$ is the fixed loading condition for the i-th experiment, and $(\mathbf{n}_i, \mathbf{s}_i)$ contains a finite set of SOH measurements where $\mathbf{s}_i^{(j)}$ is the SOH after cycle $\mathbf{n}_i^{(j)}$. The general approach to model SOH under dynamic loading is as follows:

1. Use machine learning to estimate either
   a) $\hat{f}_{\mathbf{x}}(n) \approx f_{\mathbf{x}}(n) : \mathbb{R}^5 \times [0, \infty) \to [0, 1]$, or
   b) $\widehat{\Delta f}(s, \mathbf{x}, n) \approx \Delta f(s, \mathbf{x}, n) : [0, 1] \times \mathbb{R}^5 \times [0, \infty) \to [0, 1]$.
2. Use (3) and (4) to compute the SOH for a new set of cycles $\mathbf{n}$ with dynamic loading $\mathbf{X}$.

An important aspect is that both $\hat{f}$ and $\widehat{\Delta f}$ will be subject to uncertainty due to the limited number of experiments that can be performed in practice. This is a "small data" machine learning problem where uncertainty quantification is important. We can use probabilistic machine learning to quantify uncertainty regarding $\hat{f}$ and/or $\widehat{\Delta f}$ and propagate this uncertainty to the quantity of interest $S(\mathbf{X}, \mathbf{n})$ by Monte Carlo simulation . In this study, two different approaches for how to estimate $\hat{f}_{\mathbf{x}}(n)$ with probabilistic machine learning is explored: non-parametric using Gaussian process regression (see e.g. Rasmussen & Williams (2006); Agrell & Dahl (2021)) and parametric using Bayesian neural network (see e.g. Bingham et al. (2018)).

These models have been trained on experimental data with static loading (42 experiments with 21 duplicated unique loading conditions) and applied to predict SOH on operational data with dynamic loading from one ship (Vessel C). The operational data have been converted to a table of half-cycles by cycle counting, and it is assumed that the SOH reduction from one half-cycle is half the reduction of a corresponding full cycle. This assumption might not be entirely true and adds uncertainties to the model predictions. Examples of model predictions against pairs of experiments that were held out during training are shown in Figure 6 (LOO denotes leave one out).

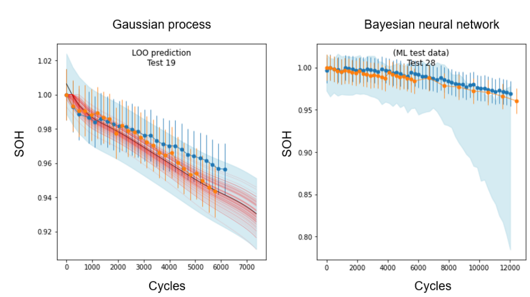

Figure 6. SOH as predicted by the Gaussian process model (left) and the Bayesian neural network (right) compared to two identical experiments (excluded from training)

The degree of uncertainty shown in Figure 6 illustrates one main challenge with this approach. For most loading scenarios (excluding those the model has been trained on), the SOH degradation can vary a lot. If such cycles are encountered, the uncertainty will be propagated and increased further in future cycles and may "blow up". This is indeed what is observed when these models are applied on the operational data; in all cases the final SOH estimate was affected by this behavior, both with the Gaussian process and the Bayesian neural network alternative.

The main conclusion from this exercise is that high-quality data and a good machine learning model is crucial for the proposed method to work, since uncertainty will accumulate in a cumulative model. Two main challenges was encountered in this study, i.e. that the training data did not cover the operational cycles well, and that there are too little degradation in the operational data in order to really evaluate model performance well.

3.1.3. Sequential and Non-Sequential Machine Learning

A set of sequential and non-sequential statistical and machine learning models were applied to the NASA randomized usage dataset based on summaries of the cumulative load profiles. I.e., features are extracted in terms of histograms or *buckets* of experienced conditions to predict capacity degradation. For further details of these analyses, reference is made to Grindheim (2022), and in the following a brief summary will be presented.

The main difference between non-sequential and sequential methods is that sequential methods take the temporal information into account, i.e. they work on time series, whereas non-sequential methods regard data as independent in time. The non-sequential methods explored in this study include linear regression without penalization (OLS) and with different types of penalization (ridge regression and lasso), gradient boosted regression trees and support vector regression (see e.g. Hastie et al. (2009) for an introduction to such methods). The sequential methods include recurrent neural networks (RNN), long short-term memory (LSTM) models, transformers and temporal convolutional neural networks (TCN), see e.g. Hochreiter & Schmidhuber (1997); Vaswani et al. (2017); Oord et al. (2016).

For the non-sequential methods, features are extracted from the random walk steps in terms of so-called *buckets* or histograms of times spent in different current profiles and with different bins of voltage and temperature. Interaction effects are included in the form of 2-dimensional buckets for current-voltage and for current-temperature as well as a 3-dimensional bucket for the time spent in different combinations of voltage, temperature and current. Additional covariates are introduced for the rest time and the temperature of the batteries as well as data from the last n steps prior to a reference cycle. The idea is that the buckets explains the cumulative degradation of the batteries, whereas the additional covariates representing the usage just prior to the reference cycles account for the stress of the battery when the reference cycles are performed. Different model alternatives that include different subsets of the histogram features – different data formats – are compared in terms of root mean square error (RMSE). The models are trained on the combined data for all cells, but where the cell to be predicted is left out.

For the sequential methods, sequential data are used as input. Two types of sequence formats are used based on the raw data: long sequences correspond to down-sampled initial time-series and short sequences include summary statistics from each random walk step. In addition, static covariates for the estimated capacity for the previous cycle and the temperature during the reference discharge are included. Some examples of predicted capacity deterioration from two of the non-sequential models and two of the sequential models are shown in Figure 7 for some of the data formats.

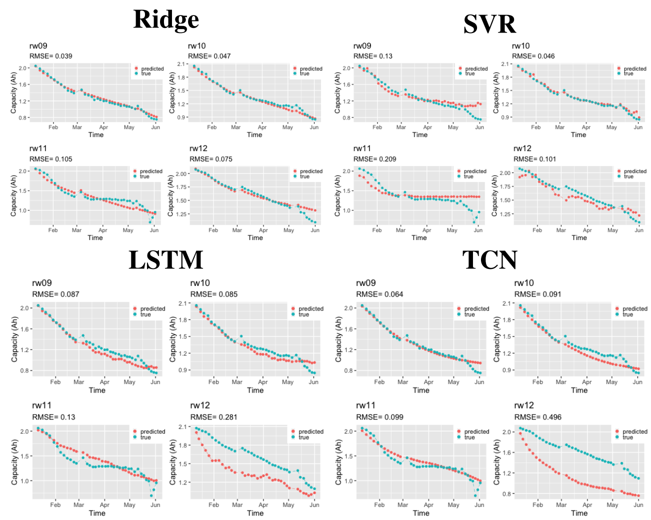

Figure 7. Predicted capacity degradation for some non-sequential methods and data format; from Grindheim (2022)

The average RMSE for the best performing models are summarized in Table 2. This summary indicate that the non-sequential penalized linear regression models – ridge regression – performs best. On second place, different models such as lasso, transformers and support vector regression, perform almost equally good on average for the four cells. These results indicate that relatively simple linear models may outperform more complicated deep learning architectures.

This exercise has illustrated that it is possible to predict battery degradation using rather simple statistical models based on buckets or histograms of operating conditions. As such it would be a cumulative damage type of model, since the complete loading history is needed in order to construct the histogram-features. Even though the data size may be significantly reduced since only summaries of the raw data are needed it is questionable how well such methods scale to large battery systems with very many battery cells, and the histograms might also be biased if there are large data-gaps

Table 2. Models ranked by lowest average RMSE score for the four batteries; from Grindheim (2022)

model	average RMSE
Ridge	0.067
Lasso	0.096
Transformer	0.097
SVR	0.099
TCN	0.108
xgboost	0.109
LSTM	0.117

in the sensor data. These are serious limitations of such types of approaches and further study would be needed in order to recommend such models for monitoring actual battery systems onboard ships.

3.1.4. Semi-Supervised Learning Approach

Access to relevant and high-quality data is a prerequisite for data-driven estimation of state of health. Some models need to be trained on a representative set of data to be applied on other battery systems, and getting access to such training data is a major challenge. For example, operational data will typically contain time series of several explanatory variables, but will very seldom include information of capacity or state of health. Typically, there will be high-resolution data of the covariates, but only few recorded values of the response from whenever the system has performed a capacity test. In such a situation, semi-supervised learning might be useful, if the limited amount of labeled data can be exploited to generate additional labels. Such an approach has been explored in Bertinelli Salucci et al. (2023), and a brief summary of this work is presented in the following.

Data from three vessels with the older battery system are used in this study and the analysis is performed on pack-level. The overall idea is to exploit the information that is in the operational data for one of the vessels to train a data-driven model to predict state of health on the two others. The data are continuous measurements of current, temperature and state of charge, and the data for Site A include results from three annual capacity tests. Hence, there are three cycles with labels. The idea is assume a time window of constant SOH around the annual test, and generate labels for other cycles within this window. This idea is illustrated by Figure 8. The pseudo-capacity corresponding to the cycles can be calculated by Coulomb counting of the incomplete cycles.

To generate labels, the data are clustered to identify groups of cycles with similar conditions. Hence, the data are filtered and clustered according to the depth of discharge (DoD), state of charge, C-rate, and temperature. Groups of similar cycles within the constant SoH-windows are regarded as reference cycles, which would get labels from the annual test results. A set of simple linear models is applied to estimate the battery capacity based on the pseudo capacity and other characteristics of the cycles. Cycles with similar characteristics are then identified in the target data, i.e., data from outside these constant SOH windows. Different cycle characteristics give rise to different groups of cycles, and SOH can be estimated independently for different groups. An illustration of the SOH estimates based on the different groups of cycles is shown in Figure 9, where different colors distinguish between different cycle groups. These estimates are for a different vessel than the one providing the training data. Final SOH estimates can be obtained by averaging the estimates from the different cycle groups, and weights are introduced reflecting the degree of uncertainty of the different groups.

An additional optional supervised learning step is suggested in Bertinelli Salucci et al. (2023), where cumulative modeling can be performed on data with the newly created SOH labels. Hence, multivariable fractional polynomial regression models are trained to model the change in SOH over time.

This method predicts a reasonable overall decreasing trend in battery capacity, although individual estimates are somewhat uncertain. However, when a vessel is used to generate training data to predict SOH of other vessels, it is questionable to what extent such a method can be used to estimate a larger extent of degradation for the target vessels compared to what the reference vessel has experienced. Hence, training data should in principle always come from the battery system with the lowest state of health, something that could be difficult to guarantee from a fleet of vessels. Notwithstanding, it is demonstrated that it is possible to extend labels of SOH from a few capacity tests to other cycles within a constant SOH window and to use this to estimate capacity degradation.

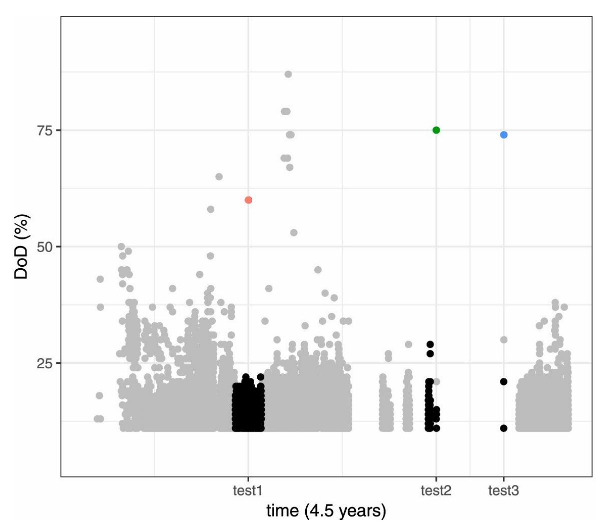

Figure 8. Data for depth of discharge for a battery pack on one of the vessels. Colored dots indicate reference cycles where SOH is observed; black dots are cycles within a window of assumed constant SOH and grey dots are cycles outside these windows. (from Bertinelli Salucci et al. (2023)).

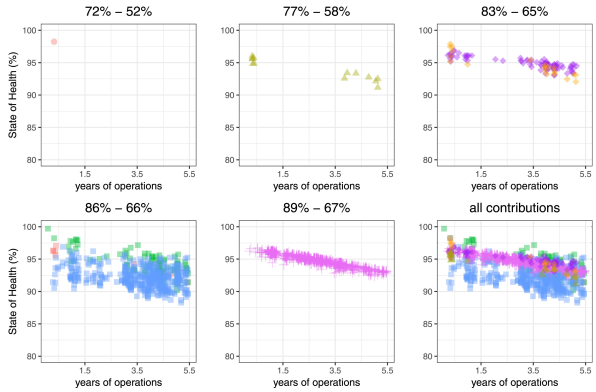

Figure 9. SOH estimates for one pack from a target vessel from different groups of cycles.

3.2. Snapshot Methods

3.2.1. Regression on Features Extracted from Charge and Discharge Curves

Results from a number of relatively simple regression models for remaining capacity based on features extracted from charge and discharge curves from the Fraunhofer laboratory data were presented in Vanem et al. (2022). This work was extended by fusing data from both Fraunhofer and Corvus lab tests in Vanem et al. (2023). In the following, a brief summary will be given.

After some initial filtering of the data, features are extracted from the charge and discharge curves from individual cycles close to the capacity measurement. Examples of extracted charge and discharge curves for an arbitrary DDE-cell are shown in Figure 10. The different colours correspond to different regular cycling periods, and it is clearly seen that these curves are changing as the battery degrades. The measured capacity preceding each period of regular cycling is also indicated in the figures, and "no = 2" indicates that these results are based on the second charge/discharge cycles in the regular cycling periods.

From the selected cycles, the mean, minimum and maximum temperature as well as the mean current are used as covariates. Moreover, the total charge throughputs between voltage ranges in steps of 0.1 V are used as additional explanatory variables. It is noted that features have only been extracted from the constant-current phase in this study. One reason for this is that these are the features deemed most likely to be found in data from battery systems in actual operation onboard ships. In total up to 44 features are collected and the overall dataset of extracted features contains 281 samples for the DDE cells and 269 samples for the DDP cells. However, it is noted that not all cells have information for all covariates. The various cells have been cycled between different voltage limits, and therefore have different subsets of the voltage-based features. Other features suggested in the literature, include features from derivative curves and from probability density functions from time spent in different voltage ranges, see e.g. Weng et al. (2013); Feng et al. (2013); Zheng et al. (2018); Jiang et al. (2020); Ibraheem et al. (2023).

A number of rather simple statistical models are employed in this study to predict the capacity of the battery cells based on snapshot features, i.e., Linear regression (Linear), Linear regression with missing covariates (Miss), Ridge regression (Ridge), Least absolute shrinkage and selection operator regression (Lasso), Multivariable fractional polynomial regression (MFP), Generalized additive models (GAM), Regression tree (RT), Random forest (RF) and Support vector regression (SVM), see Vanem et al. (2023). See also Hastie et al. (2009) for full mathematical description of the various models.

As outlined in Vanem et al. (2022, 2023), this snapshot approach yields quite good results for some of the cells, but unfortunately not for all. Examples of model predictions compared to observations are shown in Figure 11 for an arbitrary cell. The prediction performance of this approach is evaluated by calculating the average RMSE from all model alternatives for the individual cells. Results are illustrated in Figure 12, which also indicate the test parameters for the various cells. The orange markers present the average RMSE for each cell when the data from that cell is included in the training data and the red markers present the results when data from that cell have been excluded from the training. The vertical bars correspond to the voltage range the cells have been cycled between (right axis), and the two colours of each ver-

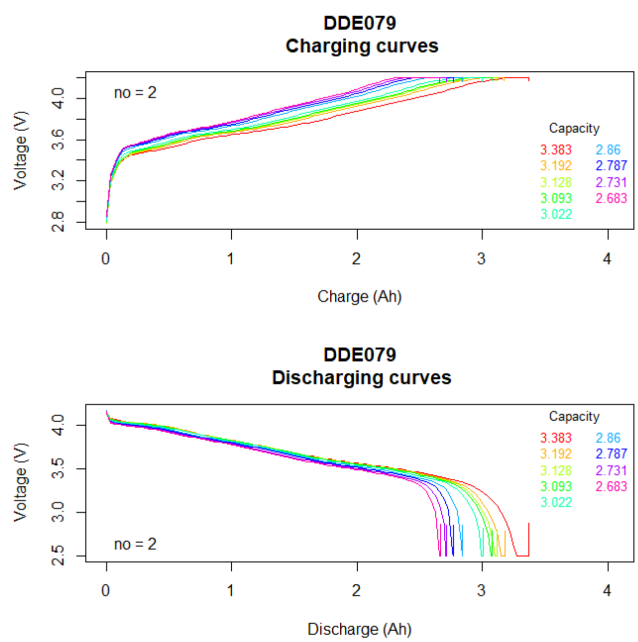

Figure 10. Extracted charge and discharge curves from the raw time series for an arbitrary DDE cell

tical bar correspond to the C-rate (leftmost colour) and temperature (rightmost color) of the different cells, respectively, as indicated by the colour legends in the plots. Evaluations of performance across predictive models are also presented in Vanem et al. (2022, 2023), without any particular model being consistently the best one.

The fact that the simple statistical models are able to yield good results for some of the cells is encouraging, but results need to be improved in order to be able to recommend such an approach from a class perspective. Presumably, more relevant training data would be needed in order to improve results, as elaborated in more detail in Vanem et al. (2022, 2023).

3.2.2. Simple Linear Model Based on Coulomb Counting

Coulomb counting is based on a fundamental relationship between integrated current and change in state of charge (SOC), than can be used to estimate total capacity of a battery. The cumulative current-time that the battery can deliver from fully charged to fully discharged state corresponds to the charge capacity of the battery (in terms of Ampere-hours (Ah)). This relationship should also be preserved during partial cycling and can be utilized in a simple linear model to estimate capacity and subsequently the state of health.

The relationship between the total capacity Q and state of charge SOC of a battery at times t_1 and t_2 is described by

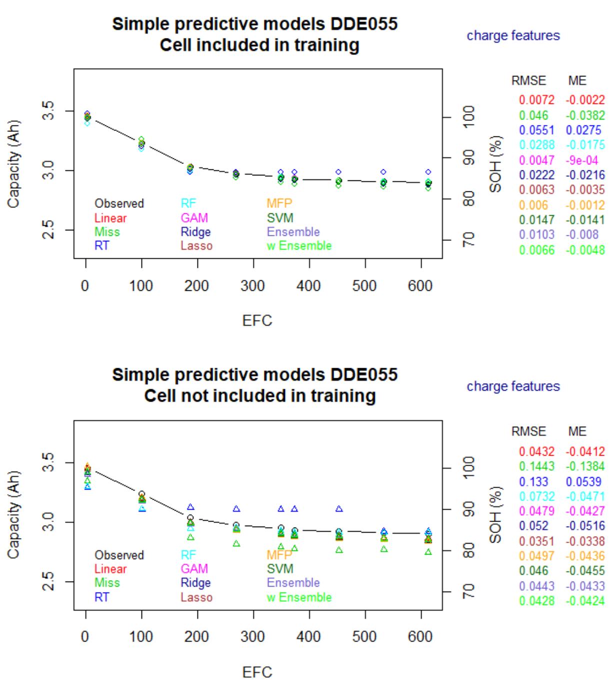

Figure 11. Example of data-driven predictions based on snapshot features from charge cycles only

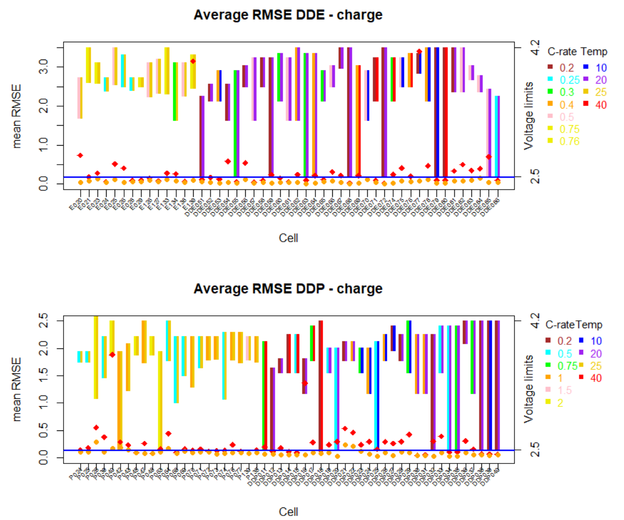

Figure 12. Average RMSE and test parameters for each cell

the following equation:

$$SOC(t_2) = SOC(t_1) + \frac{1}{Q}\int_{t_1}^{t_2} \eta I(\tau) d\tau \tag{5}$$

where $I(\tau)$ is the battery current at time τ measured in amperes, which is defined as positive when charging and negative when discharging, and η is a unitless Coulomb efficiency factor. For simplification, $\eta \approx 1$ may be assumed.

By can be rewritten as

$$\int_{t_1}^{t_2} \eta I(\tau) d\tau = Q(SOC(t_2) - SOC(t_1)) \tag{6}$$

which can be presented as a linear regression problem

$$Y = QX + \varepsilon, \tag{7}$$

where $Y = \int_{t_1}^{t_2} \eta I(\tau) d\tau$ and $X = SOC(t_2) - SOC(t_1)$, as suggested by Plett (2011). By collecting data for Y and X the regression coefficient Q, representing total capacity, can be estimated by different methods, such as ordinary least squares (OLS) and total least squares (TLS). Note that the regression model does not have an intercept; when there is no current, or when the integrated current is zero, there should also not be any change in SOC. Hence, in principle, one observation of concurrent (X, Y) should be sufficient to give an estimate of the regression coefficient. Different implementations of such a simple linear model has been applied to both laboratory and operational data, as outlined in the following.

First, a simple OLS version of this model is applied to the Fraunhofer lab data for the three different types of cells. After each capacity measurement, data are collected from the first 25 cycles and the integrated current and the change in

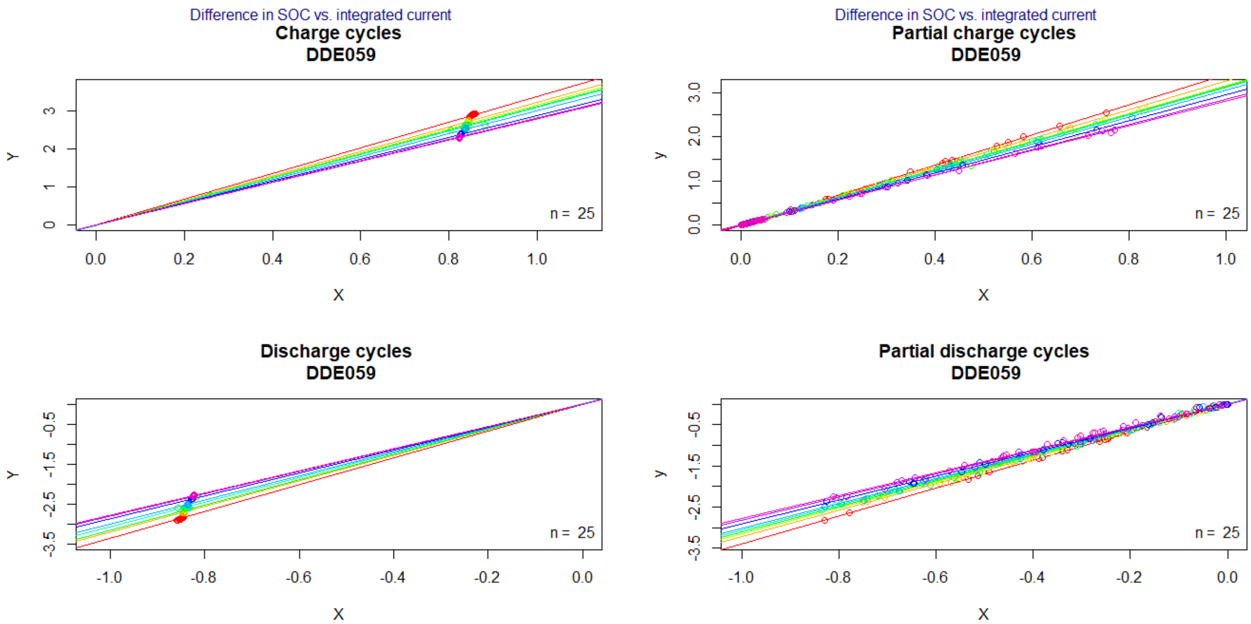

Figure 13. Applying the simple linear method to estimate SoH on one cell; Calculated X = change in SOC and Y = integrated current for the full and partial cycles together with estimates regression lines

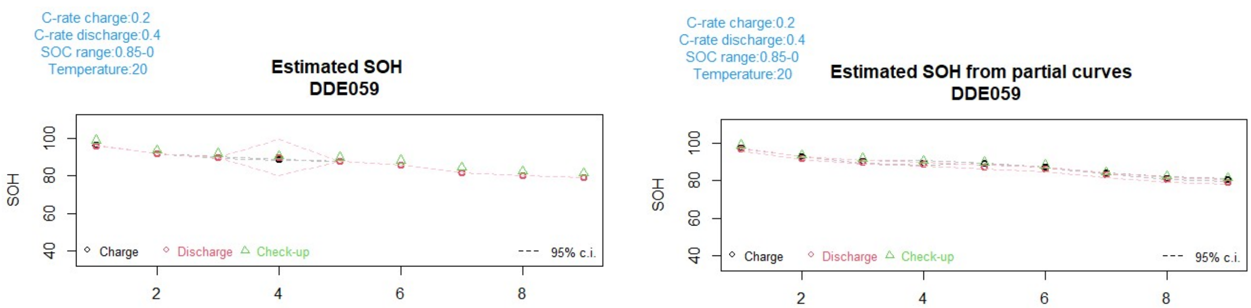

Figure 14. Estimated SOH based on full and partial cycles from the simple linear model. X-axis denote the check-ups.

SOC is calculated from the complete charge and discharge cycles, as well as from partial cycles, where random segments of the charge and discharge are extracted. Figure 13 shows the extracted data for $X = \Delta SOC$ and $Y = \int_{t_1}^{t_2} I(\tau)d\tau$ and estimated regression lines for an arbitrary cell and Figure 14 shows estimated SOH based on both full and partial cycles. Overall, the results are satisfactory and indicate that this approach can yield reasonable results for most of the cells. However, for a few cells the estimate capacity deviate notably from the measured capacity. This is most likely due to inaccuracies in the calculated SOC values.

The same method was also tried on operational data from ships in service. Different ways of filtering the data in order to get comparable capacity estimates at various stages of the batteries' life were explored and compared with annual capacity tests. Figure 15 shows estimates for four of the packs based on daily, weekly and monthly data. As can be seen, despite an overall decreasing trend, there is much variability and the uncertainty increases as the time period for the data decreases. Figure 16 shows estimated capacity based on yearly data for eight different battery packs (from two arrays). For these estimates, all data from a whole year is combined to yield and average annual capacity estimate. A decreasing trend is observed, but with some outliers, most notably in the year 2018. However, in practice one want to use this method as a snapshot method without the need to collect data from a whole year. Hence, estimates based on shorter time intervals would be preferred. Additional filtering based on C-rates, temperatures and depth of discharge were investigated, and reference is made to Kejvalova (2022a) for details.

The simple linear model can also be solved as a total least squares (TLS) problem, where uncertainties in both the change of state of charge (X) and integrated current (Y) can be taken into account. These uncertainties can be due to mea-

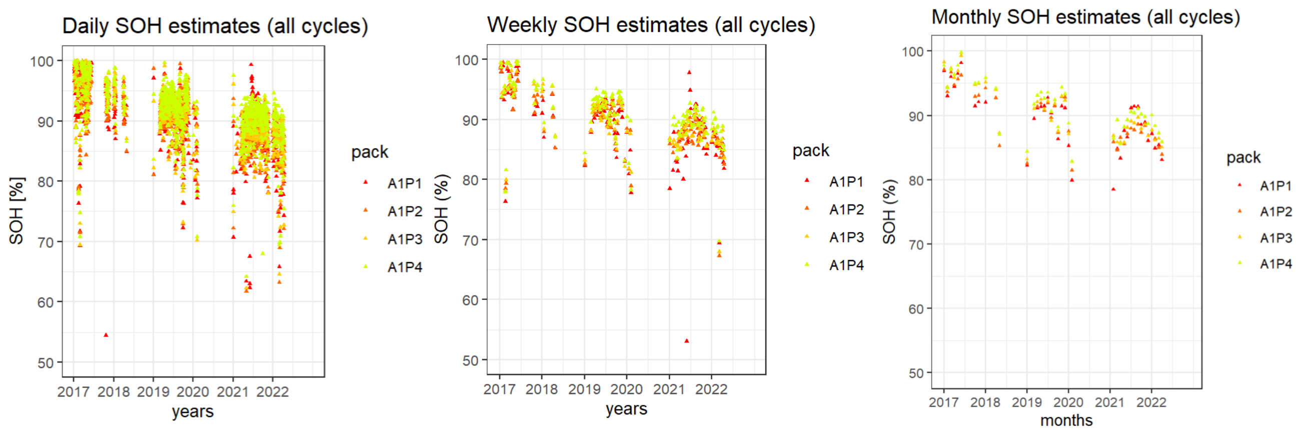

Figure 15. Daily, weekly and monthly estimates of total capacity based on simple linear model (from Kejvalova (2022a))

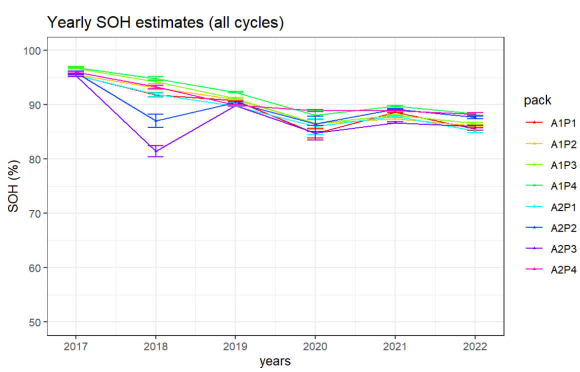

Figure 16. Yearly capacity estimates

surement error, and it is known that ignoring this in a linear regression model will tend to a bias towards zero of the regression coefficient – an attenuation bias (see Carroll et al. (2006)). The total least squares method aims at accounting for this by considering additive measurement errors in both X and Y. That is, given true values $X^\star$ and $Y^\star$, the error-prone X and Y are observed,

$$X = X^\star + \Delta X$$
$$Y = Y^\star + \Delta Y.$$

The measurement errors ΔX and ΔY can typically be assumed to be normally distributed with mean 0 and variances $\sigma^2_{\Delta X}$ and $\sigma^2_{\Delta Y}$. Hence, the linear regression model of eq. (7) can be rewritten as

$$Y^\star + \Delta Y = Q\left(X^\star + \Delta X\right) + \varepsilon \tag{8}$$

Note that the measurement error in the dependent variable can not be distinguished from the equation error ε and will not introduce a bias (only add to the variance), but the measurement error in the explanatory variable X will lead to the attenuation bias. Total least squares will seek to minimize the squares of errors in both the dependent and independent variables and hence to correct for the attenuation bias of ordinary least squares. Various TLS implementations proposed in Plett (2011) have been applied to operational data from ships in service, and details are presented in Kejvalova (2022b). The sensitivity of sensor measurement error and the influence this has on the uncertainties of X and Y is highlighted in Kejvalova (2022b), and Figure 17 illustrates the difference between OLS-estimates and TLS-estimates of total capacity based on operational data from a battery pack. Results show different results for the TLS model for different variance ratios, $\phi = \frac{\sigma^2_{\Delta x}}{\sigma^2_{\Delta y}}$; when the ratio is small, indicating that the error in X is small compared to the error in Y, then the TLS estimate is similar to the OLS estimate. However, the difference increases as the error in X grows compared to the error in Y, as expected. In all cases, the OLS estimate underestimate the regression coefficient Q (total capacity), meaning that this could be construed as a lower limit. Note that the variances (or the variance ratio) need to be known in the TLS approach.

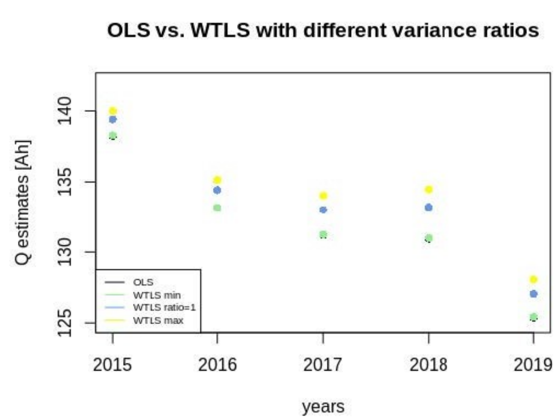

Figure 17. Difference between OLS and TLS estimates of capacity for different variance ratios (from Kejvalova (2022b))

The simple linear model in (6) has also been implemented as a Bayesian linear regression model (BLR) and applied on operational data from a ship in service. Rather than fitting the model to the data by OLS or TLS, a prior distribution for the model parameters is introduced and the likelihood is computed from the observed data to give the posterior of the model parameters using Bayes' theorem. Hence, predictive distributions rather than point estimates can be obtained, see Goldstein & Wooff (2007) for more details. This implementation of the model includes an intercept with a zero-mean Gaussian prior with small variance. The prior for the slope is based on the data in an empirical Bayes approach, see Gelman et al. (2014). An example of capacity estimates from the Bayesian linear model over time for one vessel is shown in Figure 18. It is observed that the estimated capacity initially increases, before it stabilize and tend to decrease. Such behavior may seem counter-intuitive, and needs to be carefully evaluated. However, it may also be explained by the fact that total capacity is not a fixed quantity but rather a function of the operating conditions, as discussed above. If the operational conditions change, e.g. a change in c-rate or temperature, and estimated increase in total capacity might not be unexpected.

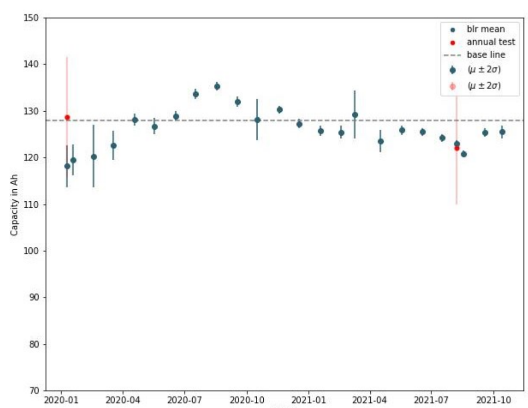

Figure 18. BLR model prediction and uncertainty over time.

In summary, the simple linear model based on Coulomb counting remains an intuitive and promising approach for condition monitoring on maritime battery systems. Some of its advantages are that there is no need for training data to train the model and that it is a snapshot method that do not rely on complete uninterrupted data-streams from the vessel. Moreover, it may equally well be applied on pack level as on cell lever, if pack SOC can be provided.

One challenge with this approach is that the total capacity of a battery is not really a fixed quantity; it is a function of several variables such as temperature, current, depth of discharge and voltage. Hence, the capacity of a particular battery for a particular time is not constant, but will vary according to how it is operated. Hence, $Q = Q(\theta)$, where θ may be a number of variables influencing the capacity (Rozas et al., 2021). It is not obvious how to account for such effects in the simple linear model. Additional covariates can be included, or careful filtering of the data can be applied prior to analysis. Another challenge with this method is that it relies heavily on state of charge, which is a derived measure and not directly observed. Ideally, the state of charge should reflect the temperature and current variations but calculating this from real data is not straightforward and there can be large uncertainties in calculated SOC values.

3.2.3. Voltage-Deviation Method

The voltage deviation method is a method for estimating battery capacity/State of Health by exploiting the relationship between decreasing capacity and increasing internal resistance and impedance. It was promoted in Yamamoto et al. (2022), where it was applied to LTO-types of batteries and reportedly performed well. In this study it is tried out on the NMC-type of batteries from the maritime battery system.

The voltage deviation method (VDM) needs to be trained in order to establish the relationship between the voltage deviation and the other features and the battery capacity/state of health. Hence, in this study, the idea is to use the lab data from the DDF-cells as training data and to apply them on operational data from some of the vessels with similar battery cells. However, it should be noted that only 6 DDF-cells have been cycled in the lab, so the amount of training data is very limited.

The voltage deviation method, is a rather simple linear regression model based on features related to the voltage deviation at certain state of charge sections and the standard deviation of charge and discharge power and mean temperature. The overall prediction model is on the form

$$
\begin{aligned}
SOH &= \alpha_1 X_1 + \alpha_2, \\
\alpha_i &= \beta_{i,1} X_2 + \beta_{i,2} X_3 + \beta_{i,3}, \quad i = 1, 2,
\end{aligned}
\tag{9}
$$

Here, X_1 denotes the voltage deviation feature (FV in Yamamoto et al. (2022)), X_2 denotes the power deviation feature (Pfv in Yamamoto et al. (2022)) and X_3 denote the average temperature (Tfv in Yamamoto et al. (2022)). The different deviation features are calculated by first dividing the data into small tiles of state of charge and calculate the standard deviation of the voltage and power, respectively, within each tile. The voltage and power deviations are then the average standard deviations over all tiles within the entire range of SoC used in the analysis. The perhaps most important feature in this model is the voltage deviation, and the idea is that when the internal resistance and impedance of the battery increases as the battery ages, the over- and underpotential occurring during charging and discharging will increase. Hence the model establishes a relationship between two effects of aging: capacity fade and resistance increase.

A SoC range of 50% - 70% was chosen since this is a range that is frequently found in the operational data. The training data are obtained from the lab data by selecting the 25 first charge and discharge cycles after a check-up and then extracting the features. An example of extracting the voltage deviation feature is shown in Figure 19. The top figure shows state of charge plotted against voltage. The different colors correspond to data from different check-up periods (25 first cycles after a check-up). It can be observed that as the battery degrades, the charge-discharge curves tend to get somewhat wider. In the bottom plot, the standard deviations of the voltage within SoC-tiles of 1% within the 50% - 70% SoC-range are shown. The final feature to use in the data-driven model is the average voltage standard deviation across all tiles. Visually, it is observed that the voltage deviation changes as the battery degrades; it increases from 0.0366 just after the initial check-up to 0.0551 for the last observation.

The linear regression model in eq. (9) is fitted to these data and yields an adjusted R^2-value of about 0.894. Then, features are extracted in the same way from the operational data. Snapshots of data from two-weeks periods are selected about 3 months apart for each cell and the VDM method is used to estimate the state of health for these periods. Preliminary results obtained with this method estimates a sudden increase on capacity after approximately one year, see Figure 20. This seems counter-intuitive, and it can be observed that this sudden shift appear at the same time as these batteries start doing fast-charging. Apparently, the models are not able to properly adjust for this change in loading. It should be noted that these results are based on only six or four cells used for training data, and lack of sufficient relevant training data can possibly explain this. Indeed, looking at the combined features from the training and the test data, as shown in Figure 21, it is clear that the power feature has a limited range in the training data

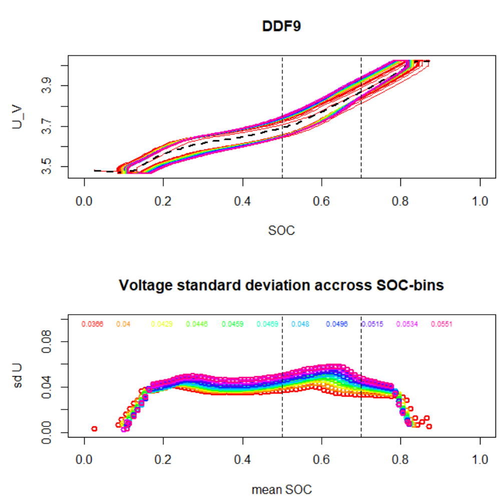

Figure 19. Extracting voltage deviation features from the training data

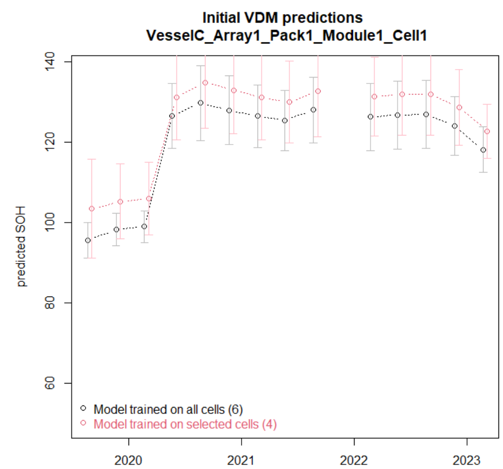

Figure 20. Predictions from preliminary analysis of the VDM method on actual field data

compared to the test data; the training data simply did not experience the same power deviation as the test data, most likely due to the fast charging. Hence, the model is not trained to account for this effect. Possibly, with an extended dataset, results would be improved.

4. DISCUSSION

Several different approaches to data-driven estimation of battery capacity have been explored in this study, with the aim of establishing a framework for monitoring state of health for class compliance. Various data sources have been exploited, including results from laboratory experiments, operational data from ships in service and other publicly available datasets. Several of the approaches yields promising results, but accurate and reliable state of health estimation remains a challenge. Some lessons learned from these investigations are discussed in the following.

Both cumulative damage models and snapshot methods have been explored, and these have different data requirements. Although initial studies have indicated that cumulative models may yield reasonable capacity estimates, these approaches are sensitive to missing data, something that must be expected to occur occasionally for ships in operation. One solution to this could be to utilize edge computing and process the data locally onboard the ship. Further studies are needed in order to investigate whether this is a feasible solution. Moreover, for very large battery systems, the amount of data becomes huge, and cumulative approaches that require all data might not scale well. One of the case studies on a cumulative method presented in this paper found that the data handling and computational cost becomes prohibitively large even for a subset of the data. Hence, snapshot methods that only need

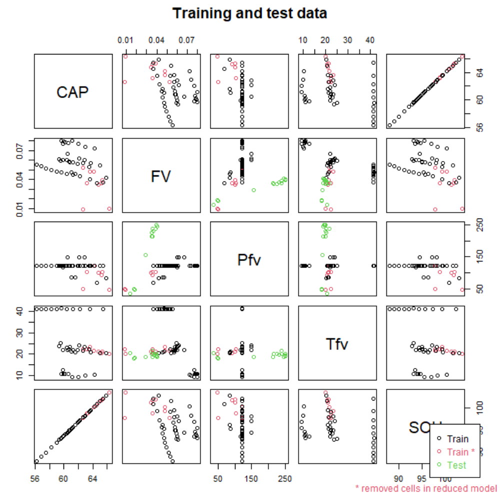

Figure 21. Pairwise scatterplots of the features for training and test data

data from certain time-intervals would be much preferred from a practical point of view.

Several snapshot approaches have been explored in this study, some of which yields promising results. One important distinction between the snapshot methods in this study is that whereas some of them require training data to train the data-driven models, the simple linear model based on Coulomb counting does not. This is believed to be a huge advantage, especially since it has been found extremely challenging to obtain sufficient relevant training data. For snapshot methods that rely on training data, carefully designed and time-consuming experiments will probably be needed in order to ensure that the trained models can predict any realistic operational condition. Hence, class requirements would need to include strict criteria for the quality, amount and representativeness of the training data. Semi-supervised methods have also been investigated in this study, with the purpose of labeling operational data for use as training data. Although this is an interesting approach, representativeness is again a challenge, and one would need to ensure that the operational data used to train the models always come from the battery systems that have experienced most degradation. This may be difficult to guarantee in practice unless complete data histories from several systems that have reached its end of life are available. Fundamentally, data-driven models are better for interpolating than extrapolating, and this will always be a challenge for models trained on a finite amount of data.

Hence, it is believed that methods similar to the simple linear model based on Coulomb counting are the most promising candidates for data-driven state of health estimation of maritime battery systems. They are snapshot methods and they do not need to be trained prior to prediction. However, one main challenge with this approach is that it relies on the state of charge, which is not measured directly but need to derived from other observed quantities. This gives rise to additional uncertainty and the accuracy of SOC would need to be verified. There are different ways of estimating SOC from observed time series of currents, voltages and temperatures, and variations in SOC estimates may introduce bias in SOH predictions. Notwithstanding, it is believed that if reliable and accurate SOC values are available, snapshot methods based on this may be useful, and verification of SOC may be included in class requirements if such approaches should be recommended. Hence, the problem of verifying SOH, which is challenging, may partly be shifted to verification of SOC, which might be easier. Alternatively, similar approaches that are based on voltage differences rather than state of charge may be attractive and this will be investigated in future research.

Different implementations of the linear model based on Coulomb counting has been explored in this study, in addition to the standard OLS method. With a Bayesian implementation, one may get a full posterior predictive distribution and one may use prior information in addition to the data if available. Moreover, it is known that the OLS solution has an attenuation bias that can be corrected by accounting for measurement errors (Plett (2011); Kejvalova (2022b)). However, in light of the other challenges and uncertainties involved, this effect is not believed to be significant, and an OLS implementation would presumably suffice. Since the attenuation bias yields a bias towards lower SOH, the OLS estimate will tend to be conservative and this can be construed as an extra safety factor rather than a serious problem.

Still, there are remaining challenges that needs to be resolved in order to ensure reliable and accurate estimates from the linear model between integrated current and state of charge. One fundamental issue is the fact that the total capacity of the battery is not a constant quantity, but will be a function of cycling conditions such as temperature, current and voltage, and it will also be affected by stresses caused by preceding conditions (it is not memoryless), see e.g. Rozas et al. (2021). It is not obvious how to best account for this in the modeling. One solution is to apply narrow filters to ensure that only data from similar conditions are compared, at the expense of significantly reduced datasets. Further research will focus on this to find optimal ways of filtering operational data. Another solution is to include additional covariates in the simple linear model, although this means that the effect of these would need to be learned from some training data. On the other hand, since this method relies on the derived quantity SOC, the effect of varying conditions may also be, at least partly, accounted for by the SOC-algorithm, Further research will focus on optimizing this approach for maritime battery systems in actual operation.

The focus has been on purely data-driven approaches in this study, and physics based or hybrid modeling approaches have not been investigated. There exist a large amount on literature on such methods, exploiting various types of equivalent circuit models or electrochemical models (see e.g. the review in Vanem, Bertinelli Salucci, et al. (2021)). It is noted that such approaches might be useful, even though they have not been investigated in this study.

Model evaluation and verification remains a big challenge and makes it difficult to arrive at conclusive recommendations. Some approaches are found to work well on laboratory data, but still fails to achieve the same accuracy and reliability when applied to actual field data. Hence, it is questionable whether lab data can be used to verify the models. Moreover, the most relevant operational data only exist for relatively new battery systems without a large extent of degradation. Hence, it is difficult to verify how the models perform when the batteries approach their end of life, when condition monitoring really becomes important. Data from older battery systems exist, but they are less relevant if the systems contain different battery cells. Furthermore, it has been observed that the quality and completeness of data from older systems are less, due to less focus on data collection and storage in the past. Hence, more robust verification of the various data-driven approaches cannot be achieved until more data with longer operational history are available from relevant battery systems.

5. SUMMARY AND CONCLUSION

This paper summarizes case studies of several approaches to data-driven estimation of state of health of lithium-ion batteries, with a focus on the needs from the maritime industry and ship classification. The various approaches, which include cumulative and snapshot methods have different challenges and advantages, and even though no conclusive recommendations on the best approach have been formulated, some important lessons have been learned. From a practical point of view, snapshot methods are much preferred over cumulative models. Moreover, approaches that do not need prior training would be much preferred over models that depend on training data. Even with semi-supervised approaches it will be challenging to obtain sufficient high-quality, relevant and representative data. Furthermore, laboratory experiments are both time consuming and costly, and lab data may not be representative for data from actual operations onboard ships. Hence, some variant of the simple linear model based on Coulomb counting are believed to be the most promising candidates for data-driven SOH verification of maritime battery systems. Further work will focus on optimizing this approach with regards to data handling and data requirements, including requirements of the required input variable state of charge.

ACKNOWLEDGMENT

This work has partly been carried out within the DDD BATMAN project, supported by MarTERA and the Research Council of Norway (project no 311445). Parts of this study have been performed in collaboration with the University of Oslo within the BIG INSIGHT SFI, supported by the Research Council of Norway (project no. 237718).

REFERENCES

Agrell, C., & Dahl, K. R. (2021). Sequential Bayesian optimal experimental design for structural reliability analysis. *Statistics and Computing*, *31*(27).

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

Bertinelli Salucci, C., Bakdi, A., Glad, I. K., Vanem, E., & De Bin, R. (2023). A novel semi-supervised learning approach for State of Health monitoring of maritime lithium-ion batteries. *Journal of Power Sources*, *556*, 232429.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., ... Goodman, N. D. (2018). Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*.

Bole, B., Kulkarni, C. S., & Daigle, M. (2014). Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use. In *Proceedings of the annual conference of the prognostics and health management society 2014.* PHM Society.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models. a modern perspective* (Second ed.). Chapman & Hall.

dos Reis, G., Strange, C., Yadav, M., & Li, S. (2021). Lithium-ion battery data and where to find it. *Energy and AI*, *5*, 100081.

Feng, X., Li, J., Ouyang, M., Lu, L., Li, J., & He, X. (2013). Using probability density function to evaluate the state of health of lithium-ion batteries. *Journal of Power Sources*, *232*, 209-218.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, B., David, Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). CRC Press.

Goldstein, M., & Wooff, D. (2007). *Bayes linear statistics: Theory and methods* (1st ed.). Wiley.

Grindheim, C. A. (2022). *Methods for battery state of health estimation* (Master's thesis). Department of Mathemathics, University of Oslo.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.

Hill, D., Agarwal, A., & Gully, B. (2015). A review of engineering and safety considerations for hybrid-power (lithium-ion) systems in offshore applications. *Oil and Gas Facilities*, *4*(3), 68-77.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Ibraheem, R., Strange, C., & dos Reis, G. (2023). Capacity and internal resistance of lithium-ion batteries: Full degradation curve prediction from voltage response at constant current at discharge. *Journal of Power Sources*, *556*, 232477.

Jiang, B., Dai, H., & Wei, X. (2020). Incremental capacity analysis based adaptive capacity estimation for lithium-ion battery considering charging condition. *Applied Energy*, *269*, 115074:1-12.

Kejvalova, A. (2022a). *Maritime battery capacity estimation* (Technical report No. 2022-0945). DNV.

Kejvalova, A. (2022b). *Total least squares estimation of maritime battery capacity* (Master's thesis). Department of Mathemathics, University of Oslo.

Liang, Q., Vanem, E., Alnes, Ø., Xue, Y., Zhang, H., Lam, J., & Bruvik, K. (2022, October). Data-driven state of health monitoring for maritime battery systems – a case study on sensor data from a ship in operation. In *Proc. of the international conference on ships and offshore structures 2022 (icsos 2022)* (p. 128-145).

Liang, Q., Vanem, E., Alnes, Ø., Xue, Y., Zhang, H., Lam, J., & Bruvik, K. (2023). Data-driven state of health monitoring for maritime battery systems – a case study on sensor data from a ship in operation. *Ships and Offshore Structures*, *Latest articles*, 1-13.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.

Plett, G. L. (2011). Recursive approximate weighted total least squares estimation of battery cell total capacity. *Journal of Power Sources*, *196*, 2319-2331.

Pop, V., Bergveld, H. J., Danilov, D., Regiten, P. P. L., & Notten, P. H. L. (2008). Battery aging process. In *Battery management systems. Accurate state-of-charge indication for battery-powered applications* (p. 111-143). Dordrecht: Springer.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

Rozas, H., Troncoso-Kurtovic, D., Ley, C. P., & Orchard, M. E. (2021). Lithium-ion battery State-of-Latent-Energy (SoLE): A fresh new look to the problem of energy autonomy prognostics in storage systems. *Journal of Energy Storage*, *40*, 102735.

Vanem, E., Alnes, Ø. Å., & Lam, J. (2021, November-December). Data-driven diagnostics and prognostics for modelling the state of health of maritime battery systems – a review. In *Proc. annual conference of the prognostics and health management society 2021 (phm 2021).*

Vanem, E., Bertinelli Salucci, C., Bakdi, A., & 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.

Vanem, E., Bruch, M., Liang, Q., Reyes Gonzalez, D. V., & Alnes, Ø. Å. (2022, October-November). A data-driven snapshot method for state of health modelling and diagnostics of maritime battery systems. In *Proc. annual conference of the prognostics and health management society 2022 (phm 2022).*

Vanem, E., Bruch, M., Liang, Q., Thorbjørnsen, K., Valøen, L. O., & Alnes, Ø. Å. (2023). Data-driven snapshot methods leveraging data fusion to estimate state of health for maritime battery systems. *Energy Storage*, *Early View*, e476.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Weng, C., Cui, Y., Sun, J., & Peng, H. (2013). On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression. *Journal of Power Sources*, 36-44.

Xue, Y., Zhou, H., Luo, Y., & Lam, J. (2022, March). Battery degradation modelling and prediction with combination of machine learning and semi-empirical methods. In *Proc. 12th International conference on Power, Energy and Electrical Engineering (CPEEE 2022)* (p. 78-85).

Yamamoto, T., Hatano, H., Maruchi, K., & Mitsumoto, K. (2022). Soundness monitoring for battery energy storage system by voltage deviation method (in Japanese). *IEEJ Transactions on Power and Energy*, *142*(1), 51-57.

Zheng, L., Zhu, J., Lu, D. D.-C., Wang, G., & He, T. (2018). Incremental capacity analysis and differential voltage analysis based state of charge and capacity estimation for lithium-ion batteries. *Energy*, *150*, 759-769.