



The battery information format (BIF): A memory-efficient standard for communicating cell usage and status for battery digital twins (BDTs)

Downloaded from: <https://research.chalmers.se>, 2026-04-18 20:27 UTC


Citation for the original published paper (version of record):

Westman, K., Aitio, A., Nilsson, V. et al (2025). The battery information format (BIF): A memory-efficient standard for communicating cell usage and status for battery digital twins (BDTs). *Journal of Power Sources*, 652: 237538-.
<http://dx.doi.org/10.1016/j.jpowsour.2025.237538>

N.B. When citing this work, cite the original published paper.



The battery information format (BIF): A memory-efficient standard for communicating cell usage and status for battery digital twins (BDTs)

Kasper Westman^{a,b,d,*} , Antti Aitio^a, Viktor Nilsson^b, Patrik Johansson^{c,d}

^a Elysia – Battery Intelligence from Fortescue Zero, Kidlington, OX5 1GB, UK, United Kingdom

^b Nortical AB, SE-41119, Göteborg, Sweden

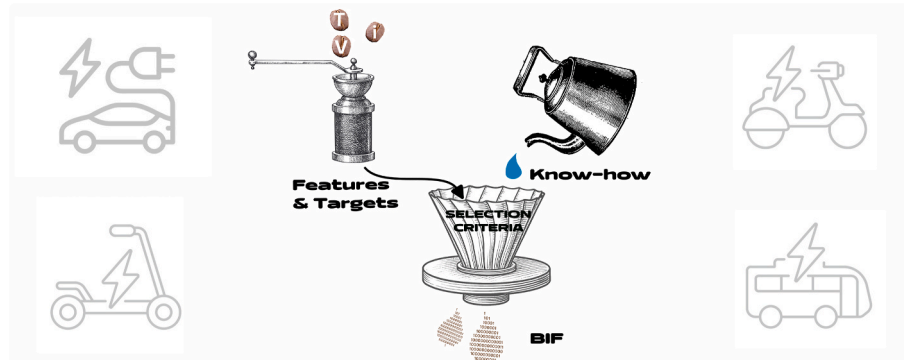
^c ALISTORE-ERI, FR CNRS 3104, Hub de l'Energie, 80039, Amiens, France

^d Department of Physics, Chalmers University of Technology, SE-41296, Göteborg, Sweden

HIGHLIGHTS

- Limited datasets curb quantitative design of field data for ageing prognostics.
- A battery information format can be used to mitigate the data shortage.
- A qualitative selection framework should form the basis for feature selection.
- Histograms are found to fulfill selection criteria.
- Harmonised format design, e.g. binning schemes, ensures data transferrability

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Battery lifetime prediction
 Battery management systems
 Battery informatics
 Battery degradation
 Battery field data

ABSTRACT

The massive expansion in lithium-ion battery (LIB) powered applications demands attention to reliability and increased lifetime, for a truly sustainable value chain. Data-driven lifetime modelling offers a route to ensure optimal usage patterns, but the lack of widely accessible LIB fleet data for training hampers benchmarking and optimization. Here we approach the lack of available training data by outlining an implementation of a battery information format (BIF). We suggest a feature-first approach, simultaneously respecting memory constraints of embedded systems and heralding a mechanistic understanding of LIB ageing. First, we examine existing feature designs. Second, we define qualitative requirements for collected battery data. From this combined, we arrive at a BIF where trade-offs can be matched to the constraints of the application. The BIF follows a histogram-based strategy, which allows to share battery state-of-health (SOH) and usage history in an interpretable, accessible, and reusable way. This BIF strategy furthermore allows training of data-driven models for predicting SOH and a vast reduction of the data footprint – by 2-4 orders of magnitude as compared to time-series data collection schemes.

This article is part of a special issue entitled: In memory of Professor Bruno Scrosati published in Journal of Power Sources.

* Corresponding author. Department of Physics, Chalmers University of Technology, SE-41296, Göteborg, Sweden

E-mail address: kasper.westman@chalmers.se (K. Westman).

<https://doi.org/10.1016/j.jpowsour.2025.237538>

Received 14 March 2025; Received in revised form 11 May 2025; Accepted 31 May 2025

Available online 21 June 2025

0378-7753/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The “electrify everything” paradigm has emerged as one of the most ambitious societal transformations of the 21st Century, including broad electrification of the transportation sector and expansion of grid-scale battery energy storage systems (BESS). It is a transformation largely enabled by the success of the lithium-ion battery (LIB) technology, implying that efficient usage of LIBs becomes increasingly important, both in the short and long term, as LIBs move from niche to mainstream applications [1]. Short-term it is important to adapt usage to delay ageing, maintaining a high battery state-of-health (SOH) and state-of-safety (SOS), thereby avoiding direct economic losses, caused by limited operating capability of the LIB. Long-term, extension of total LIB lifetime and remaining useful life (RUL), potentially by giving batteries a 2nd life, reduces both exposure to bottlenecks in raw materials and production capacity, as well as global waste flows [2,3]. Overall, minimizing the environmental impact of LIBs [4], or similar battery technologies, such as sodium-ion batteries (SIBs) [5], is essential to ensure sustainability through the life cycle, crucial for public acceptance and credibility of the transformation [6].

While LIB raw material supply chains, manufacturing stages, and assembly processes are all mature and rather well controlled, battery usage is less predictable and varies substantially across use cases, possibly affecting both the short- and long-term goals adversely. One strategy to improve this part of the battery value chain is to establish a battery digital twin (BDT), which preferably consists of three parts: *i*) a model of the battery and its constituent cells, *ii*) a record of *relevant* metadata and usage data delivered by the battery sensors, and *iii*) an information link between the physical battery and the data record [7–9]. An implemented BDT can provide better control and forecasts of battery ageing, improve battery safety, and even provide feedback on cell design choice(s) vs. usage phase matching [10–12]. Ultimately, a BDT can be used to create a more reliable 2nd life market, include batteries being traded at a higher price point, by virtue of lowering the uncertainty in prognoses of battery RUL, SOH and SOS, and thus the risk for the buyer [13,14].

Accurate and computationally efficient ways to model cell ageing (relevant to *i*) above) are plentiful [15–22], with a considerable increase in (supervised) machine learning ((S)ML) models in recent years. However, insufficient attention has been paid to *ii*) and *iii*) above, with most initiatives taken by legislators and industry consortia, mainly to approach the lack of standardized field data, e.g. via the EU battery Regulation [23] and the “Battery Pass” project [24]. These initiatives strive to increase *transparency* in the battery value chain, but do not specifically target *ageing prognostics* problems.

Consequently, a research gap exists in the design of data collection strategies and standardization of minimum viable sets of model inputs and outputs, specifically suited for *on-board collection*. In ML syntax, such inputs and outputs are also known as “*features*” and “*targets*”. A common (implicit) assumption has been that *features* and *targets* can always be extracted from high fidelity signals, originating from on-board battery voltage, current, and temperature sensors, and processed by a battery management system (BMS) [11,17,22,25,26]. In reality, and crucial for the design of BDTs, this is hardly ever the case and furthermore one reason for the apparent lack of large real-world data sets available to academia, even if commercial secrecy likely plays an even bigger role [27].

Problems with previous data set collection approaches also extends to the yearly data footprint, which can be expected to be 1–10 MB (Calculation S1). For a fleet incorporating millions-to-billions of cells, this grows into TB-to-PB/annum, which becomes prohibitively expensive for e.g. industries of mass production with slim profit margins, such as the automotive sector. The data collection problems are thus associated both with the direct costs of on-board storage as well as transmission and indirect costs, including redesign of embedded hardware and developing software for data harvesting. In addition, the cost

sensitivity and the system limitations vary heavily with application, hence there is no one-size-fits all solution. Native memory in common microcontrollers typically amounts to less than a megabyte and while expanding memory by peripheral devices is possible, it also drives cost and risk of component failure, while not being an option for legacy devices. Thus, a standard for *on-board* battery data for RUL/SOH/SOS prognoses should strive to be compact. Some typical microcontroller memory capabilities, considering various applications, and the corresponding price point are provided in Table S1.

Here we approach the challenge of on-board battery cell data collection by proposing a flexible yet standardized battery information format (BIF). The BIF targets the *logical* cell of the BDT, also known as a “supercell”, but henceforth denoted “cell” or “battery cell”, i.e. the minimal information unit in any battery, possibly composed of several *physical* cells connected in parallel. We limit the scope to the cell as the number of strategies available for aggregating data from cell to module or pack are too plentiful to be covered in detail: examples include low frequency logging of min/max voltages and temperatures, using the average cell state-of-charge (SOC) or the minimum SOC, or defining a collective of the “weakest” cells to log only them. Alternative approaches include counting special events, such as fast charges. Likewise, we limit our scope to only investigate the amount of data that in a worst-case scenario must be stored on-board, since strategies for off-boarding data are likely to also be heavily application dependent.

Focusing on the cell level BIF, we first highlight the foundations of battery cell ageing that should be incorporated into a relevant usage history. Second, we review *features* proposed elsewhere, both challenging their usefulness and claims made regarding predictive accuracy, before presenting our perspective on the qualitative aspects to be considered. Finally, we present our BIF and how it can improve BDT design and assist in reaching both the short-term and long-term goals.

2. Battery ageing fundamentals – a very brief account

Battery cells reach their end-of-function (EOF) when they can no longer provide the *useful* energy or power, collectively defining the SOH, that is required for the target application (Fig. 1, top). Useful energy and power can be expressed by cell capacity (Q) and resistance/impedance (R_{ω}), respectively: low Q and high R_{ω} diminishes useful energy, while power is mainly limited by high resistance. EOF is also reached when SOS is too low, with an exact definition of SOS remaining elusive but includes e.g. the risk of a thermal event/fire [28]. Henceforth, SOH will indicate an arbitrary combination of Q and R_{ω} and left for the application to define.

Regardless of the above, the decrease of Q and increase of R_{ω} take place through degradation modes (3rd level, Fig. 1). These are, in turn, the result of degradation mechanisms (4th level, Fig. 1), triggered by “stressors” [26], i.e. the operating condition(s) inducing a specific failure mode (5th level, Fig. 1). Several detailed accounts of ageing and the failure hierarchy in Fig. 1 exists (for readers needing a more elaborate introduction to the topic of battery degradation and ageing we refer to Refs. [29–31]) and we thus limit ourselves to a brief summary for the case of LIBs, but many of the mechanisms, such as the growth of the solid electrolyte interphase (SEI), are likely to be valid, to various extent, for many next generation battery technologies, such as SIBs, solid-state (lithium) batteries [32], calcium batteries [33], and lithium-sulfur batteries [34].

Decrease of Q can be divided into loss of (anode/cathode) active material ($LAM_{a/c}$) and loss of lithium inventory (LLI) [29]. For LAM, the mechanism involves sites for lithium insertion being lost due to structural changes at the micro or meso-scale, such as cracking upon cycling [35], $AM_{a/c}$ particles disconnecting from the electrode, or isolation of particles due to resistive surface film formation [36–40]. LLI on the other hand is caused by parasitic reactions, such as formation of the SEI at the anode surface, and filled $AM_{a/c}$ sites becoming electrically isolated within the electrode or disconnected from it [41]. LLI can further occur

due to metal plating, leading to both loss of cyclable ions, and dendrite-growth, i.e. the formation of tree-like structures that eventually risks short-circuiting the cell.

R_{ω} increases partly due to the same reasons as capacity loss; meso/microstructure evolution decreases the electronic contact and conductivity within the electrode, while growth of the SEI makes ion transport slower, both by limiting access to single $AM_{a/c}$ particles and by decreasing ion mobility/diffusivity within the porous electrode [30]. Electrolyte loss, accelerated by high temperature, poor cell design, and SEI formation, will further increase internal resistance. R_{ω} is dependent on the observation timescale/frequency, hence the ω subscript, and observably higher at long timescales, due to ohmic contributions, solid state diffusion, reaction kinetics and ion transport processes successively contributing to the voltage response of the cell, which also means that the R_{ω} timescale dependence may change as the cell ages.

Both mechanisms impacting Q and those impacting R_{ω} are triggered by stressors such as external mechanical stress, wide SOC swings, too high/low anode potentials, high cathode potential, and high temperature. The cathode/anode potential-related effects are determined *globally* by the mean SOC level [42–46] and *globally/locally* by the ion kinetics of the electrode, such effects being amplified at high currents and low temperatures [47,48]. This does not imply that raising the temperature is inherently a good mitigation strategy, since there are other kinetic effects, such as SEI dissolution, that, when progressing slowly, help keep the cell stable [49–51]. Thus, too high temperature will also impact ageing negatively [52]. It is the history of this lowest tier of stressors in Fig. 1 that should be recorded if one desires a causal relationship between historical stress and ageing.

3. Feature formats – an account

A significant body of work has accumulated that aim to predict

capacity fade, resistance increase using data driven methods, including Artificial Neural Networks (ANN) [53], Gaussian Process Regression (GPR) [54], Random Forests (RF) [55], and Support Vector Regression (SVR) [56], to name a few. These methods all depend on a feature set, which is either designed beforehand and prepared for collection by the BMS, or from off-board manipulation of raw sensor data and estimated states (such as SOC) [17,19,20,22,25,26,57–59]. With both approaches there appears to be a lack of consistency as to which are to be considered features representing the true stress history and which are to be considered health indicators (HIs), i.e., proxies for the instantaneous SOH. Several studies have designed features modelling the mapping from HIs to SOH [60–68], an endeavour that is motivated by HIs potentially being more practical to collect than capacity or resistance estimates. Below we consider the approaches used for target and feature selection, focusing on features that can be used to represent the stress history, allowing for causal modelling between usage and SOH. To benchmark data collection strategies and unless stated otherwise, we consider a *base case* where each feature requires 4 bytes of storage, since they might require pertaining large integer values over the lifetime of a battery. We also assume that 6 targets (3 health metrics + 3 associated variances) are collected, each requiring 2 bytes. The collection of features is assumed to happen 26 times per year, a time scale at which ageing becomes observable in the harshest usage modes [25].

3.1. Training targets

Forecasting SOH, RUL, and possibly SOS, involves deciding on a desired model output(s), consisting of a combination of performance metrics. Most studies have focused on the battery capacity (Fig. 2, upper green line), either predicting absolute values (Fig. 2, crosses on the green line), capacity loss (Fig. 2, ΔQ) in a predefined usage window, forecasting of the EOF or knee-point, i.e. the event where the rate of capacity

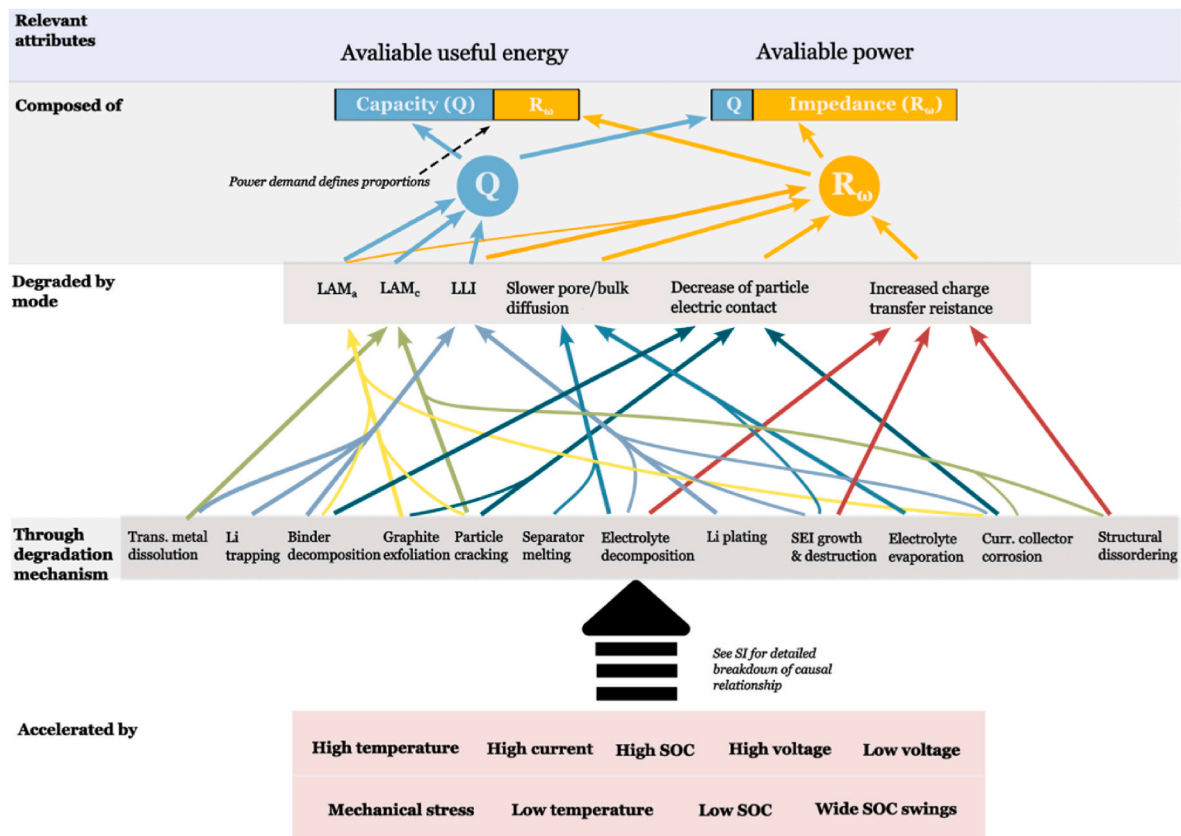


Fig. 1. The hierarchy of usage conditions and their impact on degradation mechanisms, degradation modes, and subsequently on battery performance metrics. Heavily inspired by C. Birkl [30]. For links between usage and failure modes see Fig. S1.

loss increases, resulting in an increasingly rapid capacity fade (topmost circle, Fig. 2) [17–19,21,58,69,70]. A few studies have rather focused on the resistance, predicted pointwise, knee point, and the EOF (Fig. 2, purple line) [20,71]. Common for all targets is that their instantaneous estimation constitutes a challenging problem that requires the targets to be rigorously defined [72].

3.2. Training inputs – usage history

Once the training target has been defined, the most essential task in a SML problem is to design the input, consisting of a feature set of arbitrary dimensionality. The four most common strategies for selection and curation of features are: derivation from the incremental capacity curve, semi-empirical (point) features, histograms, and event-counting (Table 1).

3.2.1. Incremental capacity-curve features

Introduced as a ML feature by Severson et al. in their 2019 investigation of data-driven prediction of ageing in 124 LFP cells (hereafter denoted as the “Stanford dataset”), the incremental (dis)charge capacity vs. voltage curve (ICV-curve) has been used as a base for features in several publications [11,25,69,73]. Most of these use the difference between curves at equivalent cycles x and y , $\Delta Q_{x-y}(V)$, and calculate statistical measures such as variance, kurtosis, min/max from the difference, and use these as features in ageing models [11,69]. The predictive accuracy reached by employing this method on the Stanford dataset and smaller complementary sets, reaches a root mean square error (RMSE, SI DEF 2) for RUL prediction of 10 %, as a function of cycles or charge throughput. Assuming storage of 1000 voltage and capacity points at 16-bit resolution, a total data footprint of 4 kB is required to enable storage of the curve, meaning 104 kB of annual storage.

The ICV-curve approach lends itself readily to lab testing with well-defined cell cycles, but in real life scenarios, where loading patterns tend to be dynamic and can vary substantially between cells, capturing a full (dis)charge curve at a given cycle will likely not be feasible. Since the idea of a cycle is poorly defined, as SOC start- and endpoints can vary between (dis)charging events, some studies have instead employed energy or capacity throughput as the relevant metric to compare with capacity fade. This, however, does not change if the underlying data generation process is still based on cycles, as in the Stanford dataset. The Stanford dataset also suffers from a narrow feature space mainly containing stressful charging conditions, likely masking relevant usage modes with more narrow SOC windows and lower charging currents. Consequently, the results (including RMSE) from studies employing the Stanford dataset data should only be considered relevant in the context of the experimental set-up, and not assumed to be generalizable to other

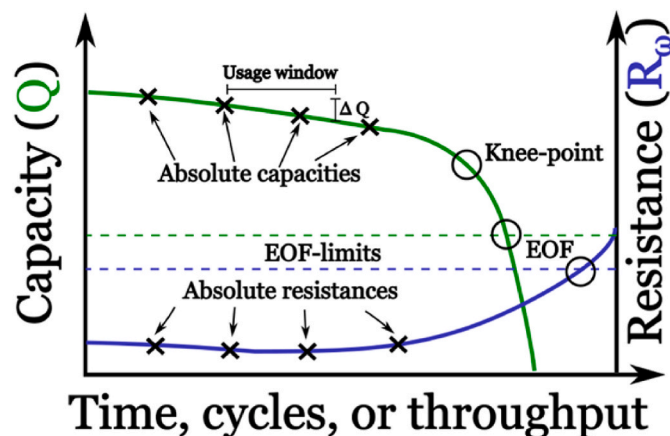


Fig. 2. Capacity- and resistance-based prediction targets.

use cases.

Finally, the modelling strategy using $\Delta Q_{x-y}(V)$ -derived features encodes a mapping from the ICV-curve to usage pattern severity, and from severity to RUL, leading to a form of data leakage [69,79]. Such an approach is likely to fail whenever the usage pattern changes – more of a rule than an exception in a real-life use-case.

3.2.2. Pointwise semi-empirical features

Features based on phenomenological understanding on battery cell ageing include equivalent cycle counting, calendar time, square root of time, charge throughput, mean SOC, mean temperature in a usage time window, to name but a few [80]. They have been employed in studies using both the Stanford dataset and the randomized cycling of 28 LCO cells recorded by NASA (hereafter denoted as the “NASA dataset”) and serve as the most lightweight feature category observed in literature, including the upcoming European Union battery passport legislation [15,24,58,74–76]. A data footprint in the range of 1 kB per annum is achieved for 10 features, reaching a RMSE, in total capacity predictions, in the range 1–4 %, when the capacity is predicted from beginning-of-life to 50 % nominal capacity. The semi-empirical features represent the usage history of the battery cell, albeit in a coarse-grained manner, and could thus be associated with ageing. However, it is still an open question if any subsets have adequate representativity to address ageing in a large share of cell use cases. As already pointed out, the validation data used in these studies is limited or represents relatively homogenous conditions, not providing enough evidence for the generality of any specific set. Again, whenever a “cycle” is used as a feature, it is likely to be protocol dependent concept that cannot easily be transferred across different usage modes and applications.

3.2.3. Histograms

Still informed by the physics governing battery cell ageing but addressing the problem of representativity, *histograms* are another category of features explored. So far, multidimensional histograms with 1–3 axis dimensions and time as counting dimension have been used, resulting in so-called “residence-time counting” histograms [17,19,22,26,57,74]. For example, Greenbank et al. used the Stanford data set together with a GPR model and conducted automatic feature selection from a set of features derived from hours spent between the percentiles of current, voltage, temperature, power, and absolute power and current [17]. Similarly, both Zhang et al. and von Bülow et al. used 1D-3D histograms derived from a proprietary PHEV dataset (Zhang [19], mainly with data on pack level) and the Stanford dataset to train ANN and GPR models to predict the capacity fade in an upcoming window of cell operation. The axis dimensions included instantaneous current, 3-min RMS current, cell temperature, and either DOD or SOC. In both cases, separate histograms were used for charge, discharge, resting (von Bülow [26]), driving or parked (Zhang [19]). The authors employed either residence time (von Bülow [26]), or a combination of residence time, energy throughput, and density/normalized residence time (Zhang [19]) as count dimension. All reviewed studies basing their model input on histograms, reported a median RMSE error of 0.4–2 percentage units for the prediction of capacity, where the upper limit is conservative and based on Zhang et al. using a percentage/relative error version of RMSE [19], noting that the capacity range over which the prediction is not always specified even though it is expected to impact the RMSE value significantly. Likewise, RUL and knee point were predicted with an RMSE of 1.3 and 2.6 percent, respectively, where again EOF was defined as capacity declining to 80 % [17].

The data footprint for histogram-based features will naturally be larger than for the previous two feature categories. For example, a single collection event of the histograms used by Zhang et al. requires 388 bytes of storage and results in an annual footprint of ca. 10 kB. In the end, the data footprint for histograms will depend on the binning scheme employed, even if several studies have turned to post-processing to decrease the number of features by removing highly correlated bins,

making model training more feasible. This, however, requires the full scope of data already available and cannot be used to judge the data footprint at the point of collection. We suggest that none of the studies conducted so far can argue for *one* optimal binning scheme, nor collection frequency, even if the levels of error seen in the studies are below 2 % for prediction of pointwise capacity. This is because also in the histogram case, the size of datasets used for validation are still too small and narrow in scope to make a judgement of feasibility. Indeed, since histograms tend to expand the total number of features, the datasets used for training and validation need to be larger to avoid overfitting, assure generalizability and increase model precision. As an example of when global model validity suffers from pre-processing can be seen in the paper of Greenbank et al., observing that features relating to temperature are excluded because of their high correlation with power. Experience and understanding of battery cell ageing, however, tells us that temperature should be included as a feature of its own merit [81,82]. Expanding the test protocol to control for a set of elevated temperatures or varying the temperature while keeping current constant would likely result in temperature being decorrelated from power, forcing it to be included.

3.2.4. Event counting and snapshots

A final category of features for battery cell health prognostics are those based on event counting, meaning one or more signals to be collected whenever a pre-defined event occurs. Pointwise counters, for example indicating the number of fast-charging sessions or snapshots [78,83] have been proposed and are included in the data long-list for the EU battery passport [24]. As an example, Johnsson used a proprietary dataset generated by Volvo C30 test vehicles [78], generating features based on history of pack level aggregates collected at each time the battery pack transitioned between operational modes. The aggregates consisted of the operational mode, the total lifetime of the battery pack, the duration of the operational mode, as well as the mean, max, and average of bus voltage, DoD, power, battery pack temperature and ambient temperature. This results in an approximate feature footprint of 72 bytes per collection, but an annual footprint is not possible to define. However, assuming three state changes per day gives an estimate of *ca.* 79 kB/annum. Since the target SOH in this data set was noisy, only limited predictive power was presented, questioning the validity of features and single cell health prediction.

3.2.5. Summary of the current state of feature sets

Several attempts have been made to model the SOH trajectory of batteries with various feature sets and data-driven methods. Most of the studies predicting capacity(change) or EOF demonstrate an RMSE of 1–10 %, with studies utilizing more features giving lower RMSE and marginally higher precision. Unfortunately, the correlation between data footprint and predictive power is far from clear, likely relating to a lack of solid evidence in the datasets. This implies that it is impossible to draw conclusions about the generalizability and prediction power of these features for useful energy and power prediction in real-life applications.

4. The BIF

To support the development of a BDT that can improve battery lifetime efficiency, we proceed to define a BIF (“battery information format”, as a courtesy to the reader). The BIF must have the generality to fit several different applications in predicting useful energy and power and possibly SOS. First, we present specific performance indicators and

requirements to be used to evaluate the merits of a BIF. Second, we suggest targets and features to be included in the BIF. Finally, detailed BIF design considerations are presented. An assumption made henceforth is that the BIF data are connected to the battery cell via a unique identifier.

4.1. Selection criteria - BIF requirements

The *data footprint* clearly constitutes a design requirement for a BIF. Here we define this and four other requirements—*interpretability*, *accessibility*, *predictive power*, and *reusability* – in detail. We acknowledge that these qualities semantically overlap with the principles of FAIR research data [84], but are here refined to the specific domain of battery ageing prognostics.

4.1.1. Data footprint (DF)

By data footprint we consider how many bytes of data that need to be logged annually for a given feature set. A low footprint is desired, reducing transmission, processing, and storage of data, each driving cost. Furthermore, fewer redundant variables in the feature set improves training speed and robustness by avoiding overfitting. Reusing our base case (Section 3), 6 targets and 20 features require 2.392 kB of data per year and cell. For a Nissan Leaf model year 2019, having a 96s3p configuration, *i.e.* 96 logical cells, this would amount to *ca.* 230 kB/year.

4.1.2. Interpretability (I)

By interpretability we mean the degree of which features allow for a physical interpretation, making it possible for humans to understand them and propose alternate scenarios still relevant for the application [26]. This is valid also for black-box ML, where it, apart from helping the human suggest other relevant input scenarios, can help determine if the model is conflicting with expert knowledge. As an example, the mean SOC is easily interpretable as the mean state of lithiation, while aggregated statistics, such as higher order moments of residence time histograms, are harder to interpret. The interpretability should be as high as possible, acknowledging that it often will come at the expense of data footprint.

4.1.3. Accessibility (A)

By accessibility we mean the ease with which a feature can be collected from an engineering perspective. Accessibility can be impaired by difficulty to observe the quantity of interest or when very specialized and expensive sensors are needed, *e.g.* local internal heating in the cell is likely a stressor but hard to observe in field applications. Accessibility also denotes the risk of artificial variability introduced due to variation in definitions or implementation of collection software. As an example, aggregated features or those relying on a state estimate dependent on the battery manufacturer’s implementation of the underlying algorithm. Such variability hides “true” estimates and lowers accessibility. Accessible features and targets are unanimous and can easily be read out with low risk of artificial variability.

4.1.4. Predictive power (PP)

Predictive power means how well a feature set allows predictions to be made. The main prediction targets and modelling strategies are mentioned above in section 3, but there are other less explored targets worth predicting as well, including fast/slow dynamics resistance, available useful energy, and SOS. A BIF should yield predictions with high accuracy and reliable error bounds for as many of these targets as possible.

4.1.5. Reusability (R)

Reusability is defined as the ease with which features collected from one application could be transferred to another application, especially useful if the same (type of) cells have been used. However, data could partially be reused for initialization of ageing models also for other cases, given that it has been collected in such a way that the datasets can be reused – either directly or through proper resampling.

4.2. Targets and input feature selection

The BIF needs to specify both targets and features, as exemplified above. Relevant prediction targets for battery cell operation include cell capacity and cell resistance, analogous to available useful energy and available power. For the former, we narrow the scope to the full discharge capacity at 0.1C, 20 °C, acknowledging that an arbitrary interpretation of capacity can lead to vastly different results if the estimation method and conditions are varied [72]. To predict power capability both for short and for longer current pulses we suggest using both 1 s and 10 s DC resistance (R_{1s} , R_{10s}) at 20 °C, 50 % SOC and 1C discharge (SI, DEF 1). Furthermore, each target should be reported together with an estimated standard deviation, to judge the quality of the target value estimate. Table 1

Moving on to feature selection, we acknowledge a risk of making a completely arbitrary choice and thus apply our selection criteria on the categories observed in literature and give them a *Low*, *Medium*, or *High* score in each criterion (Table 2). Incremental capacity curve features score low on both data footprint and accessibility since it requires the whole incremental capacity curve to be captured every time the feature is to be recorded. Furthermore, the interpretability score is also low since it is not clear how to modify the curve, nor its mean, variance, and kurtosis, to propose an alternate scenario relating to a change in operational pattern of the battery. Finally, we argue that even the R score is low due to the issue of data leakage and the likelihood that these features might not contain any information about usage history.

As for the other feature categories, they rank similarly in DF with significant flexibility in how much data to collect but with event counting scoring slightly lower in accessibility due to it being non-trivial to define an event, which will introduce artificial variability into the feature. Also, for event counting, the reusability score will be low, since for some applications certain events never occur, e.g. a "driving" event never happens in a BESS. For histograms and semi-empirical features, we assume they are equal in all aspects considered. However, histograms in general and residence time-counting histograms in particular, seem to offer a greater flexibility and can capture a richer picture of the battery usage history, while putting upper bound on the data footprint. Furthermore, they are attractive from a causal standpoint, since they contain the necessary components to describe kinetics for parasitic reactions; both information about the time spent in a certain condition, and the nature of the rate-determining conditions (i,V,T) can be contained.

With the reasoning above and since 2D histograms of residence times have been successfully used to predict ageing, albeit with limited real-world validation data (section 3.2), such histograms are suggested as a starting point for a BIF. We suggest histograms including voltage, temperature, current, SOC, and a 30 s moving average of current ($i_{MA\ 30s}$) to be used (Table 4). The latter, defined in the SI, captures sustained current, something otherwise lost when forming histograms, and resembles the 3-min RMS-feature used by Zhang et al. [19] but accounts for directionality of the current (charge and discharge are separated) making it more suitable to capture local overcurrent and over-/underpotential conditions. We also require the discharge energy

throughput and charge throughput to be included to promote a fair benchmarking of the ageing vs. usage.

Even if limitations in available datasets make it impossible to empirically judge the predictive power of the features outlined here, we can still repeat the exercise from the feature category selection and rank the targets and features in descending order of accessibility (Tables 3 and 4). Targets are by necessity less accessible than features since they require an implementation of parameter estimators [85]. High frequency voltage and current signals with good noise characteristics, R_{1s} and $\sigma_{R_{1s}}$, likely render high accessibility, while Q and σ_Q , that tend to depend on a large enough SOC window traversed in a short enough time, likely are less accessible. R_{10s} and $\sigma_{R_{10s}}$ will have moderate to low accessibility, dependent on access to clean current pulses. From a fleet-of-cells perspective, however, low accessibility of targets need not be an issue if they are still recorded by part of the fleet that is large enough.

Training features will be relatively easier to access than targets, especially those based on raw sensor data. Single-cell measurements of temperature should be considered as having high accessibility, that will change to a moderate or low accessibility if temperature is estimated or extrapolated from a distant sensor. An effective moving average can be implemented in different ways, thus the accessibility for $i_{MA\ 30s}$ is lower, but it could still be directly extracted from the current. The lowest accessibility belongs to histograms containing SOC, as its estimation can be implemented in several ways; Coulomb counting, voltage mapping, Kalman filter, and furthermore also will differ between manufacturers and as often only a pack-level SOC is available [85].

4.3. BIF design

Several schemes have recently been proposed for the alignment of battery data, e.g. the "Battery Data Genome" [86] and a unified ontology [87], and the BIF aligns with and extends these. Based on the suggested targets and features, we now proceed to define units and binning strategy. For reusability, observed quantities should be of similar magnitudes and bins should satisfy necessary overlap criteria so that fine-grained and course-grained histograms can be compared.

First, we consider units (Tables 5 and 1st row), requiring that current is expressed as a directional C-rate, i.e., current is normalized by the mark-plate capacity of the given (super) cell. Voltage (in volt), and temperature (in °C) stay within the same order of magnitude/range meaning no normalization is needed. For histogram count dimension we suggest using minutes to allow for capture of rare conditions.

Second, there is a need to establish an alignment of the histogram bin edges (Fig. 3, dashed line), requiring the specification of where to put at least one edge. Without alignment there is a risk of offset between data collected from different cells, impacting the reusability. Alignment points should be chosen such that they cover most battery cell use cases since all data formats will need to include this.

Third, for the data format to allow for histogram down-sampling, we suggest using a base partitioning scheme for the bins – from here on called *base layout*, and rules for refinement. A base layout (Fig. 3, middle bin layout) is the coarsest bin size allowed, and it needs to allow for sufficient data minimization while being detailed enough to exploit transferability of data. For simplicity, we propose using an evenly spaced binning, starting from the alignment point and continuing as far as is required by the application at hand. We note, however, that the base layout could be designed with non-uniform bin partitions but leave this for future work to explore. Allowing for design freedom, any application should be allowed to split the base grid into further partitions, but to maintain transferability, any partitioning scheme should

Table 1
Selected data features used to predict capacity fade and RUL.

Input feature space	Original collection freq.	Model	RMSE _{ΔQ} (2 σ) ^a [%]	RMSE _Q (2 σ) ^a [%]	RMSE _{EOF} (2 σ) ^a [%]	Annual data footprint ^b [kB]	Reference (Dataset)	Comment
Variance of the curve ΔQ _{2,100} (V)	Once	LR	N/A	N/A	11 (N/A)	104	Severson et al. [25,73]	
Q-throughput Δt	26 times before EOF	GPR	2.0 (29)	4.3 (29)	N/A	0.5	Richardson et al. [74] ([75])	
Cycle number	Once per cycle	GPR	N/A	1 (4.8)	3 (>100)	0.4	Richardson et al. [58] ([75–77])	Multi-output GPR when knowing the full capacity fade of another cell.
V _{1,2} V _{2,3} P _{1,2} i _{3,4} t Notation X _{a,b} indicating time spent between a:th and b:th quartile of quantity X	Sporadic	GPR	0.13 (<0.19)	0.83 (<1.6)	<1.3 (<3.6)	16	Greenbank et al. [17] ([11, 25])	Data footprint based on full histograms w/50 bins per histogram, as in Ref. [17]. Source also includes prediction of knee point.
Battery state ∈ {charging, idle, in operation} For each state (change) record: Lifetime of battery State duration min, max, average of: bus voltage, DoD, Power, Battery T, Ambient T	On battery state change.	NN	N/A	1.7 (9.9)	N/A	79	Johnsson [78]	Not as accurate to estimate a data footprint. Here we assume 3 state transitions per day during a year.
Case1: 2D histograms of residence time in i vs. T (7x8 bins) SOC vs. T (7x13 bins) and SOC rain-flow counting histograms (10x12 bins) Case 2: Q-throughput Last measured/estimated Q Min and Max values of SOC and T	Every drive-cycle	SVR	N/A	Case 1: <11 Case 2: N/A	N/A	Case 1: 28 Case 2: 0.83	Nuhic et al. [57]	Approximate RMSE based on Figure 10 in Ref. [57].
2D residence time histograms [hours]: i vs. T @ charge (3x14 bins) i vs. T @ discharge (4x14 bins) i vs. SOC @ charge (3x5 bins) i vs. SOC @ discharge (4x5 bins) T vs. SOC @ charge (14x5) T vs. SOC @ discharge (14x5) T vs. SOC @ rest (14x5)	Every n cycle, where n ∈ {25,50,100}	NN	0.15	N/A	N/A	36	v. Bülow et al. [26] ([25])	Assuming the “Coarse” resolution scenario
1 D histograms counting E _{throughput} [kWh]: T @ driving (10 bins) T @ charging (10 bins) SOC @ driving (10 bins) SOC @ charging (10 bins) 1 D residence time histograms [min]: T @ parking (10 bins) SOC @ parking (10 bins) 1 D probability histograms: i _{RMS3min} discharge (14 bins)	Sporadic	RF, GPR, NN,SVR	N/A	2.11 (5) ^c	N/A	10	Zhang et al. [19] ([19,25, 75])	Real fleet PHEV data as one test scenario but [25,76] where also used, albeit with other features.

(continued on next page)

Table 1 (continued)

Input feature space	Original collection freq.	Model	RMSE _{ΔQ} (2 σ) ^a [%]	RMSE _Q (2 σ) ^a [%]	RMSE _{EOF} (2 σ) ^a [%]	Annual data footprint ^b [kB]	Reference (Dataset)	Comment
$i_{RMS3min}$ charge (9 bins) DOD (10 bins) Counters Δt Accumulated parking time $E_{throughput}$ in driving $E_{throughput}$ in charging								

^a Values in parenthesis denote the standard deviation in the prediction.

^b Estimated with the base case outlined in Section 3.

^c RMS percentage error (RMSPE) – not directly comparable to absolute RMS errors for SOH.

Table 2
Scoring of feature categories with respect to design requirements.

Feature Category	DF score	I score	A score	PP score	R score
Incremental capacity curve-derived	Low	Low	Low	Unknown	Low
Event counting	Medium	High	Low	Unknown	Low
Histograms	High	High	Medium	Unknown	High
Semi-empirical	High	High	Medium	Unknown	High

Table 3
Accessibility of targets.

Target	Accessibility
R_{1s}	High
$\sigma_{R_{1s}}$	High
R_{10s}	Low
$\sigma_{R_{10s}}$	Low-Moderate
Q	Low-Moderate
σ_Q	Low

Table 4
Histograms suggested to be used as BIF features.

Target property/Count dimension	Histogram Axes	Stressors captured	Accessibility
Q, R, R_{10s} /Time	i vs. V	High (cathode) voltage/low anode voltage, high current	High
	i vs. T	High current at low temperature, high current	High (if access to singles cell temperature measurement)
	V vs. T	High (cathode) voltage/low anode voltage, high temperature	High (if access to singles cell temperature measurement)
	V vs. $i_{MA 30s}$	High (cathode) voltage/low anode voltage, high sustained current	Moderate
	SOC vs. T	High/Low mean lithiation, high temperature	Low-Moderate
	SOC vs. i	High/Low mean lithiation, high current	Low-Moderate
	SOC vs. $i_{MA 30s}$	High/Low mean lithiation, high sustained current	Low-Moderate
Fair benchmarking/None	Energy and charge throughput	None	Moderate

only split existing bins in two (Fig. 3, lower green bin layouts). A divisor of 2 is selected as it impacts the data footprint the least.

Fourth, the last bin on each side of the histogram should be required to capture any remaining data “falling off the edge” (Fig. 3, edges of bin layouts). Finally, the collection frequency is specified, here suggested to

Table 5
Suggested BIF design hyperparameters.

Quantity /Dimension	Voltage/V	i/C -rate	$i_{MA 30s}/C$ -rate	SOC /%	T /°C
Alignment point	3.0	0	0	50	15
Base layout spacing	0.5	1	1	25	10
Bin aggregation dimension	Minutes				
Collection frequency	Biweekly				
Bin refinement divisor	2				

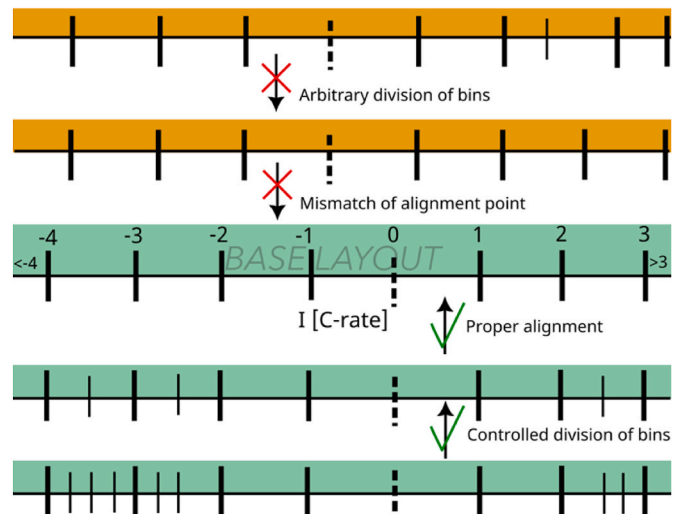


Fig. 3. Proposed histogram bin construction rules and impact on reusability if (not) following them.

be an aggressive base-case of biweekly collection, leaving the exact collection frequency as a design parameter, but requiring it to be an integer multiple of the biweekly interval. By doing so, transferability is achieved. In Table 5 we summarize our full proposal of BIF design hyperparameters.

5. Conclusions and future outlook

In conclusion, we have shown how it is possible to design a BIF based on 2D histograms for ageing prediction that respects requirements for data footprint, interpretability, accessibility, predictive power and transferability. We further argue that at this point in time, any such format should rely as much on heuristic and mechanistic understanding of battery failure modes, and engineering limitations and memory constraints of embedded systems, as it does on the limited empirical evidence available in the academic literature. In this light, our BIF is most likely not fully optimized, but *good enough* as a starting point to standardise collection of battery field data for lifetime prognostics to a degree that provides a realistic opportunity for many OEMs to implement it. By concretely voicing such realistic principles of a dataset that the scientific community would benefit from, we take one step closer towards widely available field data.

With the histogram-based approach it is instructive to compare the cell data footprint of the BIF to that of time-series data, previously concluded to be in the range of 1–10 MB/annum. Assuming that all targets and features in Tables 3 and 4 are used and that we have on average 6 bins per axis dimension, the storage requirement for one instance of the cell BIF ends up being $(7 * 6 * 6 * 4 \text{ bytes} + 2 * 4 \text{ bytes}) + (6 * 2 \text{ bytes}) = 1 \text{ kB}$, where the first expression within parenthesis relates to the storage requirements of all features and the second relates to targets. For the yearly data footprint this means histograms requiring 26 kB/annum, reducing the data footprint from time series data by at least 2 orders of magnitude. This has the promise of confining world battery fleet data to TB, avoiding the PB domain.

To reiterate, the BIF presented herein is a starting point for optimization of histogram axis dimensions, binning, and collection frequency, to balance the trade-off between data footprint vs. predictive power. However, we stress that already now, an OEM or system operator that have not yet devised a data collection strategy, could/should consider using (a subset) of the BIF features and targets presented above to not lose valuable insights. Future work should also include the evaluation of valid features and targets for prognosing SOS. It will also be relevant to develop computationally fast and efficient strategies for generating the residence time counting histograms.

CRedit authorship contribution statement

Kasper Westman: Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Antti Aitio:** Writing – review & editing, Supervision, Conceptualization. **Viktor Nilsson:** Writing – review & editing, Supervision. **Patrik Johansson:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

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.

Acknowledgements

This work was supported by the Kamprad Family Foundation within the Research and Innovation Program (Grant No. 20200165), the RE: source Strategic Innovation Programme (Swedish Energy Agency Grant No. P2024-00627), the Swedish Research Council (Grant No. 2021-

00613), and VINNOVA/BatteriesSweden (BASE) (Grant No. 2019-00064). Finally, with respect to this special issue, we would like to convey to all readers how much Prof. Scrosati has meant to us, directly and indirectly, not the least by his many inspiring visits to Chalmers, where he also was awarded *doctor honoris causa* in 2008.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpowsour.2025.237538>.

Data availability

No data was used for the research described in the article.

References

- [1] IEA, Tracking clean energy progress 2023, Paris, <https://www.iea.org/reports/tracking-clean-energy-progress-2023>, 2023. (Accessed 23 January 2024).
- [2] J. Nicholson, P. Patel, A. Lewis, De-bottlenecking the battery materials midstream, EY Web (2023). https://www.ey.com/en_se/strategy/how-europe-can-unblock-the-midstream-battery-materials-bottleneck. (Accessed 16 July 2024).
- [3] L. Boer, A. Pescatori, M. Stuermer, C. Baumeister, C. Bogmans, K. Brooks, D. Gielen, B. Gilbert, R. Kellogg, L. Kilian, C. Koch, H. Lütkepohl, P. Mattner, L. Menkhoff, E. Pritfi, I. Petrella, G. Primiceri, B. Schumann, G. Schwerhoff, N. Valckx, Energy transition metals: bottleneck for net-zero emissions? J. Eur. Econ. Assoc. 2023 (2023) 1–30, <https://doi.org/10.1093/jeaa/jvad039>.
- [4] M. Chordia, A. Nordelöf, L.A.W. Ellingsen, Environmental life cycle implications of upscaling lithium-ion battery production, Int. J. Life Cycle Assess. 26 (2021) 2024–2039, <https://doi.org/10.1007/s11367-021-01976-0>.
- [5] S. Wickerts, R. Arvidsson, A. Nordelöf, M. Svanström, P. Johansson, Prospective life cycle assessment of sodium-ion batteries made from abundant elements, J. Ind. Ecol. 28 (2024) 116–129, <https://doi.org/10.1111/jiec.13452>.
- [6] A. Nurdiawati, T.K. Agrawal, Creating a circular EV battery value chain: end-of-life strategies and future perspective, Resour. Conserv. Recycl. 185 (2022) 106484, <https://doi.org/10.1016/j.resconrec.2022.106484>.
- [7] A. Rasheed, O. San, T. Kvamsdal, Digital twin: values, challenges and enablers from a modeling perspective, IEEE Access 8 (2020) 21980–22012, <https://doi.org/10.1109/access.2020.2970143>.
- [8] A. Sancarlos, M. Cameron, A. Abel, E. Cueto, J.L. Duval, F. Chinesta, From ROM of Electrochemistry to AI-Based Battery Digital and Hybrid Twin, Springer Netherlands, 2020, <https://doi.org/10.1007/s11831-020-09404-6>.
- [9] B. Wu, W.D. Widanage, S. Yang, X. Liu, Battery digital twins: perspectives on the fusion of models, data and artificial intelligence for smart battery management systems, Energy and AI 1 (2020) 100016, <https://doi.org/10.1016/j.egyai.2020.100016>.
- [10] W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte, D.U. Sauer, Digital twin for battery systems: cloud battery management system with online state-of-charge and state-of-health estimation, J. Energy Storage 30 (2020) 101557, <https://doi.org/10.1016/j.est.2020.101557>.
- [11] 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, P.K. Herring, M. Aykol, S.J. Harris, R.D. Braatz, S. Ermon, W.C. Chueh, Closed-loop optimization of fast-charging protocols for batteries with machine learning, Nature 578 (2020) 397–402, <https://doi.org/10.1038/s41586-020-1994-5>, 7795 578 (2020).
- [12] J.M. Reniers, G. Mulder, D.A. Howey, Unlocking extra value from grid batteries using advanced models, J. Power Sources 487 (2021) 229355, <https://doi.org/10.1016/j.jpowsour.2020.229355>.
- [13] M. Baumann, S. Rohr, M. Lienkamp, Cloud-connected battery management for decision making on second-life of electric vehicle batteries, in: 2018 13th International Conference on Ecological Vehicles and Renewable Energies, EVER 2018, 2018, pp. 1–6, <https://doi.org/10.1109/ever.2018.8362355>.
- [14] D. Beverungen, B. Klör, S. Bräuer, M. Monhof, Will They Die Another Day? A Decision Support Perspective on Reusing Electric Vehicle Batteries. ECIS 2015 research-in-progress Papers, 2015. https://aisel.aisnet.org/ecis2015_rip/36. (Accessed 22 November 2022).
- [15] M. Lucu, E. Martinez-Laserna, I. Gandiaga, K. Liu, H. Camblong, W.D. Widanage, J. Marco, Data-driven nonparametric Li-ion battery ageing model aiming at learning from real operation data - part B: cycling operation, J. Energy Storage 30 (2020) 101410, <https://doi.org/10.1016/j.est.2020.101410>.
- [16] M. Dubarry, D. Beck, Big data training data for artificial intelligence-based Li-ion diagnosis and prognosis, J. Power Sources 479 (2020) 228806, <https://doi.org/10.1016/j.jpowsour.2020.228806>.
- [17] G. Samuel, D. Howey, Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life, IEEE Trans. Ind. Inf. (2021), <https://doi.org/10.1109/tii.2021.3106593>.
- [18] A. Peter, M.B. Alexander, P. Ferran, Brosa, D. Philipp, R. Goncalo, dos, D. Matthieu, G. Paul, G. Richard, G. Samuel, H. David, L. Ouyang, K. Edwin, P. Yuliya, S. Abhishek, S. Shashank, S. Anna, G.S. Valentin, Review—"Knees" in lithium-ion battery aging trajectories, J. Electrochem. Soc. 169 (2022), <https://doi.org/10.1149/1945-7111/ac6d13>.

- [19] Y. Zhang, T. Wik, J. Bergström, M. Pecht, C. Zou, A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data, *J. Power Sources* 526 (2022) 231110, <https://doi.org/10.1016/j.jpowsour.2022.231110>.
- [20] A. Aitio, D.A. Howey, Predicting battery end of life from solar off-grid system field data using machine learning, *Joule* (2021), <https://doi.org/10.1016/j.joule.2021.11.006>.
- [21] V. Sulzer, P. Mohat, A. Aitio, S. Lee, Y.T. Yeh, F. Steinbacher, M.U. Khan, J.W. Lee, J.B. Siegel, A.G. Stefanopoulou, D.A. Howey, The challenge and opportunity of battery lifetime prediction from field data, *Joule* 5 (2021) 1934–1955, <https://doi.org/10.1016/j.joule.2021.06.005>.
- [22] S. Greenbank, D.A. Howey, Piecewise-linear modelling with automated feature selection for Li-ion battery end-of-life prognosis, *Mech. Syst. Signal Process.* 184 (2023) 109612, <https://doi.org/10.1016/j.ymssp.2022.109612>.
- [23] European Parliament, REGULATION (EU) 2023/1542 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 12 July 2023 Concerning Batteries and Waste Batteries, Amending Directive 2008/98/EC and Regulation (EU) 2019/1020 and Repealing Directive 2006/66/EC, European Parliament, 2023. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32023R1542>. (Accessed 29 January 2024).
- [24] Battery Pass.eu, Resources - the battery pass, (n.d.). <https://thebatterypass.eu/resources/> (accessed July 16, 2024).
- [25] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fragedakis, M.Z. Bazant, S.J. Harris, W.C. Chueh, R. D. Braatz, Data-driven prediction of battery cycle life before capacity degradation, *Nat. Energy* 4 (2019) 383–391, <https://doi.org/10.1038/s41560-019-0356-8>, 5 4 (2019).
- [26] F. von Bülow, J. Mentz, T. Meisen, State of health forecasting of Lithium-ion batteries applicable in real-world operational conditions, *J. Energy Storage* 44 (2021) 103439, <https://doi.org/10.1016/j.est.2021.103439>.
- [27] G. dos Reis, C. Strange, M. Yadav, S. Li, Lithium-ion battery data and where to find it, *Energy and AI* 5 (2021) 100081, <https://doi.org/10.1016/j.egyai.2021.100081>.
- [28] Y. Preger, L. Torres-Castro, T. Rauhala, J. Jeevarajan, Perspective—on the safety of aged lithium-ion batteries, *J. Electrochem. Soc.* 169 (2022) 030507, <https://doi.org/10.1149/1945-7111/ac53cc>.
- [29] J. Vetter, P. Novák, M.R.R. Wagner, C. Veit, K.-C.C. Möller, J.O.O. Besenhard, M. Winter, M. Wohlfahrt-Mehrens, C. Vogler, A. Hammouche, Ageing mechanisms in lithium-ion batteries, *J. Power Sources* 147 (2005) 269–281, <https://doi.org/10.1016/j.jpowsour.2005.01.006>.
- [30] C.R. Birkl, Diagnosis and Prognosis of Degradation in Lithium-Ion Batteries, University of Oxford, 2017. https://ora.ox.ac.uk/objects/uuid:7d8c8b9c-1469-4209-9995-5871fc908b54/download_file?file_format=pdf&safe_filename=DPhil_thesis_Christoph_Birkl_final_submission.pdf.
- [31] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, D. Riu, A review on lithium-ion battery ageing mechanisms and estimations for automotive applications, *J. Power Sources* 241 (2013) 680–689, <https://doi.org/10.1016/j.jpowsour.2013.05.040>.
- [32] M. Pasta, D. Armstrong, Z.L. Brown, J. Bu, M.R. Castell, P. Chen, A. Cocks, S. A. Corr, E.J. Cussen, E. Darnbrough, V. Deshpande, C. Doerrer, M.S. Dyer, H. El-Shinawi, N. Fleck, P. Grant, G.L. Gregory, C. Grovenor, L.J. Hardwick, J.T.S. Irvine, H.J. Lee, G. Li, E. Liberti, I. McClelland, C. Monroe, P.D. Nellist, P.R. Shearing, E. Shoko, W. Song, D.S. Jolly, C.I. Thomas, S.J. Turrell, M. Vestli, C.K. Williams, Y. Zhou, P.G. Bruce, Roadmap on solid-state batteries, *J. Phys.: Energy (Calg.)* 2 (2020) 032008, <https://doi.org/10.1088/2515-7655/ab95f4> (2020).
- [33] M.R. Palacin, P. Johansson, R. Dominko, B. Dlugatch, D. Aurbach, Z. Li, M. Fichtner, O. Luzanin, J. Bitenc, Z. Wei, C. Glaser, J. Janek, A. Fernández-Barquín, A.R. Mainar, O. Leonet, I. Urdampilleta, J.A. Blázquez, D.S. Tchitchevkova, A. Ponrouch, P. Canepa, G.S. Gautam, R.S.R.G. Casilda, C.S. Martínez-Cisneros, N. U. Torres, A. Varez, J.Y. Sanchez, K.V. Kravchik, M.V. Kovalenko, A.A. Teck, H. Shiel, I.E.L. Stephens, M.P. Ryan, E. Zemlyanushin, S. Dsoke, R. Grieco, N. Patil, R. Marcilla, X. Gao, C.J. Carmalt, G. He, M.M. Titirici, Roadmap on multivalent batteries, *J. Phys.: Energy (Calg.)* 6 (2024) 031501, <https://doi.org/10.1088/2515-7655/ad34fc>.
- [34] J.B. Robinson, K. Xi, R.V. Kumar, A.C. Ferrari, H. Au, M.M. Titirici, A.P. Puerto, A. Kucernak, S.D.S. Fitch, N.G. Araez, Z.L. Brown, M. Pasta, L. Furness, A.J. Kibler, D.A. Walsh, L.R. Johnson, C. Holc, G.N. Newton, N.R. Champness, F. Markoulidis, C. Crean, R.C.T. Slade, E.I. Andritsos, Q. Cai, S. Babar, T. Zhang, C. Lekakou, N. Kulkarni, A.J.E. Rettie, R. Jervis, M. Cornish, M. Marinescu, G. Offer, Z. Li, L. Bird, C.P. Grey, M. Chhowalla, D. Di Lecce, R.E. Owen, T.S. Miller, D.J.L. Brett, S. Liatard, D. Ainsworth, P.R. Shearing, Roadmap on lithium sulfur batteries, *J. Phys.: Energy (Calg.)* 3 (2021) 031501, <https://doi.org/10.1088/2515-7655/abd9a> (2021).
- [35] U. Kasavajjula, C. Wang, A.J. Appleby, Nano- and bulk-silicon-based insertion anodes for lithium-ion secondary cells, *J. Power Sources* 163 (2007) 1003–1039, <https://doi.org/10.1016/j.jpowsour.2006.09.084>.
- [36] C.R. Birkl, M.R. Roberts, E. McTurk, P.G. Bruce, D.A. Howey, Degradation diagnostics for lithium ion cells, *J. Power Sources* 341 (2017) 373–386, <https://doi.org/10.1016/j.jpowsour.2016.12.011>.
- [37] M.S.E. Houache, C.H. Yim, Z. Karkar, Y. Abu-Lebdeh, On the Current and future outlook of battery chemistries for electric vehicles-mini review, *Batteries* 8 (2022) 70, <https://doi.org/10.3390/batteries8070070>, 70 8 (2022).
- [38] S.Y. Lai, K.D. Knudsen, B.T. Sejersted, A. Ulvestad, J.P. Mæhlen, A.Y. Kuposov, Silicon nanoparticle ensembles for lithium-ion batteries elucidated by small-angle neutron scattering, *ACS Appl. Energy Mater.* 2 (2019) 3220–3227, <https://doi.org/10.1021/acsaem.9b00071>.
- [39] X. Wang, F. Fan, J. Wang, H. Wang, S. Tao, A. Yang, Y. Liu, H. Beng Chew, S. X. Mao, T. Zhu, S. Xia, High damage tolerance of electrochemically lithiated silicon, *Nat. Commun.* 6 (1) (2015) 1–7, <https://doi.org/10.1038/ncomms9417>, 6 (2015).
- [40] A. Schiele, B. Breitung, A. Mazilkin, S. Schweidler, J. Janek, S. Gumbel, S. Fleischmann, E. Burakowska-Meise, H. Sommer, T. Brezesinski, Silicon nanoparticles with a polymer-derived carbon shell for improved lithium-ion batteries: investigation into volume expansion, gas evolution, and particle fracture, *ACS Omega* 3 (2018) 16706–16713, <https://doi.org/10.1021/acsomega.8b02541>.
- [41] F. Liu, R. Xu, Y. Wu, D.T. Boyle, A. Yang, J. Xu, Y. Zhu, Y. Ye, Z. Yu, Z. Zhang, X. Xiao, W. Huang, H. Wang, H. Chen, Y. Cui, Dynamic spatial progression of isolated lithium during battery operations, *Nature* (2021) 659–663, <https://doi.org/10.1038/s41586-021-04168-w>, 2021 600:7890 600.
- [42] W.M. Dose, J.K. Morzy, A. Mahadevegowda, C. Ducati, C.P. Grey, M.F.L. De Volder, The influence of electrochemical cycling protocols on capacity loss in nickel-rich lithium-ion batteries, *J Mater Chem A Mater* 9 (2021) 23582–23596, <https://doi.org/10.1039/D1TA06324C>.
- [43] F. Friedrich, B. Strehle, A.T.S. Freiberg, K. Kleiner, S.J. Day, C. Erk, M. Piana, H. A. Gasteiger, Capacity fading mechanisms of NCM-811 cathodes in lithium-ion batteries studied by X-ray diffraction and other diagnostics, *J. Electrochem. Soc.* 166 (2019) A3760, <https://doi.org/10.1149/2.0821915JES>.
- [44] M. Merz, B. Ying, P. Nagel, S. Schuppler, K. Kleiner, Reversible and irreversible redox processes in Li-Rich layered oxides, *Chem. Mater.* 33 (2021) 9534–9545, <https://doi.org/10.1021/acs.chemmater.1c02573>.
- [45] Z. Ahaliabadeh, X. Kong, E. Fedorovskaya, T. Kallio, Extensive comparison of doping and coating strategies for Ni-rich positive electrode materials, *J. Power Sources* 540 (2022) 231633, <https://doi.org/10.1016/j.jpowsour.2022.231633>.
- [46] A. Mikheenkova, O. Gustafsson, C. Misiewicz, W.R. Brant, M. Hahlin, M.J. Lacey, Resolving high potential structural deterioration in Ni-rich layered cathode materials for lithium-ion batteries operando, *J. Energy Storage* 57 (2023) 106211, <https://doi.org/10.1016/j.est.2022.106211>.
- [47] I. Traskunov, A. Latz, Localized fluctuations of electrochemical properties in porous electrodes of lithium-ion batteries: beyond porous electrode theory, *Electrochim. Acta* 379 (2021) 138144, <https://doi.org/10.1016/j.electacta.2021.138144>.
- [48] A. Mistry, T. Heenan, K. Smith, P. Shearing, P.P. Mukherjee, Asphericity can cause nonuniform lithium intercalation in battery active particles, *ACS Energy Lett.* 7 (2022) 1871–1879, <https://doi.org/10.1021/acsenenergylett.2c00870>.
- [49] M. Moshkovich, Y. Gofer, D. Aurbach, Investigation of the electrochemical windows of aprotic alkali Metal (Li, Na, K) salt solutions, *J. Electrochem. Soc.* 148 (2001) E155, <https://doi.org/10.1149/1.1357316>.
- [50] R. Mogensen, D. Brandell, R. Younesi, Solubility of the Solid Electrolyte Interphase (SEI) in sodium ion batteries, *ACS Energy Lett.* 1 (2016) 1173–1178, <https://doi.org/10.1021/acsenenergylett.6b00491>.
- [51] CATL unveils its latest breakthrough technology by releasing its first generation of sodium-ion batteries. <https://www.catl.com/en/news/665.html>, 2021. (Accessed 16 July 2024).
- [52] A. Blyr, C. Sigala, G. Amattucci, D. Guyomard, Y. Chabre, J.-M. Tarascon, Self-discharge of LiMn₂O₄/C li-ion cells in their discharged state: understanding by means of three-electrode measurements, *J. Electrochem. Soc.* 145 (1998) 194–209, <https://doi.org/10.1149/1.1838235>.
- [53] D.J.C. MacKay, Information Theory, Inference and Learning Algorithms, Cambridge University Press, 2003. https://books.google.se/books?id=AKuMj4PN_EMC.
- [54] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, 2006, in: <http://www.gaussianprocess.org/gpml/>. (Accessed 22 June 2021).
- [55] T.K. Ho, Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, 1995, pp. 278–282, <https://doi.org/10.1109/icdar.1995.598994>.
- [56] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (1995) 273–297, <https://doi.org/10.1007/bf00994018>.
- [57] 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, <https://doi.org/10.1016/j.jpowsour.2012.11.146>.
- [58] 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, <https://doi.org/10.1016/j.jpowsour.2017.05.004>.
- [59] S. Voronov, M. Krysanter, E. Frisk, Predictive maintenance of lead-acid batteries with sparse vehicle operational data, *Int J Progn Health Manag* 11 (2020) 1–17, <https://doi.org/10.36001/ijphm.2020.v11i1.2608>.
- [60] Y. Li, D.I. Stroe, Y. Cheng, H. Sheng, X. Sui, R. Teodorescu, On the feature selection for battery state of health estimation based on charging–discharging profiles, *J. Energy Storage* 33 (2021) 102122, <https://doi.org/10.1016/j.est.2020.102122>.
- [61] X. Hu, Y. Che, X. Lin, Z. Deng, Health prognosis for electric vehicle battery packs: a data-driven approach, *IEEE ASME Trans. Mechatron.* 25 (2020) 2622–2632, <https://doi.org/10.1109/tmech.2020.2986364>.
- [62] S.S. Mansouri, P. Karvelis, G. Georgoulas, G. Nikolakopoulos, Remaining useful battery life prediction for UAVs based on machine learning, *IFAC-PapersOnline* 50 (2017) 4727–4732, <https://doi.org/10.1016/j.ifacol.2017.08.863>.
- [63] C. Hu, G. Jain, C. Schmidt, C. Strief, M. Sullivan, Online estimation of lithium-ion battery capacity using sparse Bayesian learning, *J. Power Sources* 289 (2015) 105–113, <https://doi.org/10.1016/j.jpowsour.2015.04.166>.
- [64] M. Berecibar, M. Garmendia, I. Gandiaga, J. Crego, I. Villarreal, State of health estimation algorithm of LiFePO₄ battery packs based on differential voltage curves

- for battery management system application, *Energy (Calg.)* 103 (2016) 784–796, <https://doi.org/10.1016/j.energy.2016.02.163>.
- [65] M. Berecibar, F. Devriendt, M. Dubarry, I. Villarreal, N. Omar, W. Verbeke, J. Van Mierlo, Online state of health estimation on NMC cells based on predictive analytics, *J. Power Sources* 320 (2016) 239–250, <https://doi.org/10.1016/j.jpowsour.2016.04.109>.
- [66] K.H. Tseng, J.W. Liang, W. Chang, S.C. Huang, Regression models using fully discharged voltage and internal resistance for state of health estimation of lithium-ion batteries, *Energies* 8 (2015) 2889–2907, <https://doi.org/10.3390/en8042889>, 8 (2015) 2889–2907.
- [67] D. Yang, Y. Wang, R. Pan, R. Chen, Z. Chen, A neural network based state-of-health estimation of lithium-ion battery in electric vehicles, *Energy Proc.* 105 (2017) 2059–2064, <https://doi.org/10.1016/j.egypro.2017.03.583>.
- [68] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, L. Zhang, Remaining useful life prediction for lithium-ion battery: a deep learning approach, *IEEE Access* 6 (2018) 50587–50598, <https://doi.org/10.1109/access.2018.2858856>.
- [69] V. Sulzer, P. Mohtat, S. Lee, J.B. Siegel, A.G. Stefanopoulou, Promise and challenges of a data-driven approach for battery lifetime prognostics, in: *Proceedings of the American Control Conference 2021-May, 2021*, pp. 4427–4433, <https://doi.org/10.23919/ACC50511.2021.9483312>.
- [70] W. Li, H. Zhang, B. van Vlijmen, P. Dechent, D.U. Sauer, Forecasting battery capacity and power degradation with multi-task learning, *Energy Storage Mater.* 53 (2022) 453–466, <https://doi.org/10.1016/j.ensm.2022.09.013>.
- [71] R. Ibraheem, C. Strange, G. dos Reis, Capacity and internal resistance of lithium-ion batteries: full degradation curve prediction from voltage response at constant current at discharge, *J. Power Sources* 556 (2023) 232477, <https://doi.org/10.1016/j.jpowsour.2022.232477>.
- [72] S. Lee, P. Mohtat, J.B. Siegel, A.G. Stefanopoulou, J.W. Lee, T.K. Lee, Estimation error bound of battery electrode parameters with limited data window, *IEEE Trans. Ind. Inf.* 16 (2020) 3376–3386, <https://doi.org/10.1109/TII.2019.2952066>.
- [73] P.M. Attia, K.A. Severson, J.D. Witmer, Statistical learning for accurate and interpretable battery lifetime prediction, *J. Electrochem. Soc.* 168 (2021) 090547, <https://doi.org/10.1149/1945-7111/ac2704>.
- [74] R.R. Richardson, M.A. Osborne, D.A. Howey, Battery health prediction under generalized conditions using a Gaussian process transition model, *J. Energy Storage* 23 (2019) 320–328, <https://doi.org/10.1016/j.est.2019.03.022>.
- [75] C. Bole, B. Kulkarni, M. Daigle, Randomized Battery Usage Data Set, NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2014. <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>. (Accessed 16 July 2024).
- [76] B. Saha, K. Goebel, Battery data set, NASA ames prognostics data repository. <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>, 2007. (Accessed 16 July 2024).
- [77] D. Liu, J. Pang, J. Zhou, Y. Peng, M. Pecht, Prognostics for state of health estimation of lithium-ion batteries based on combination gaussian process functional regression, *Microelectron. Reliab.* 53 (2013) 832–839, <https://doi.org/10.1016/j.microrel.2013.03.010>.
- [78] H. Johnsson, A Neural Network Approach to Absolute state-of-health Estimation in Electric Vehicles Battery Degradation Study Based on Fleet Data, Chalmers University of Technology, Master Thesis, 2018. <https://odr.chalmers.se/handle/20.500.12380/255593>. (Accessed 14 January 2022).
- [79] A. Geslin, B. van Vlijmen, X. Cui, A. Bhargava, P.A. Asinger, R.D. Braatz, W. C. Chueh, Selecting the appropriate features in battery lifetime predictions, *Joule* 7 (2023) 1956–1965, <https://doi.org/10.1016/j.joule.2023.07.021>.
- [80] V. Kumtepli, H.C. Hesse, M. Schimpe, A. Tripathi, Y. Wang, A. Jossen, Energy arbitrage optimization with battery storage: 3D-MILP for electro-thermal performance and semi-empirical aging models, *IEEE Access* 8 (2020) 204325–204341, <https://doi.org/10.1109/access.2020.3035504>.
- [81] J. Groot, *State-Of-Health Estimation of Li-ion Batteries: Ageing Models*, Doctoral Thesis, Chalmers University of Technology, 2014.
- [82] E. Wikner, T. Thiringer, Extending battery lifetime by avoiding high SOC, *Applied Sciences (Switzerland)* 8 (2018), <https://doi.org/10.3390/app8101825>.
- [83] S. Rohr, S. Müller, M. Baumann, M. Kerler, F. Ebert, D. Kaden, M. Lienkamp, Quantifying uncertainties in reusing lithium-ion batteries from electric vehicles, *Procedia Manuf.* 8 (2017) 603–610, <https://doi.org/10.1016/j.promfg.2017.02.077>.
- [84] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M. E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR guiding principles for scientific data management and stewardship, *Sci. Data* 3 (1) (2016) 1–9, <https://doi.org/10.1038/sdata.2016.18>, 3 (2016).
- [85] G.L. Plett, *Battery Management Systems – Battery Modeling*, Artech House, 2015.
- [86] L. Ward, S. Babinec, E.J. Dufek, D.A. Howey, V. Viswanathan, M. Aykol, D.A. Beck, B. Blaiszk, B.-R. Chen, G. Crabtree, V. de Angelis, P. Dechent, M. Dubarry, E.E. Eggleton, D.P. Finegan, I. Foster, C. Gopal, P. Herring, V.W. Hu, N.H. Paulson, Y. Preger, D.U. Sauer, K. Smith, S. Snyder, S. Sripad, T.R. Tanim, L. Tao, Principles of the battery data genome, *Joule* 6 (2022) 2253–2271, <https://doi.org/10.1016/j.joule.2022.08.008>.
- [87] S. Clark, F.L. Bleken, S. Stier, E. Flores, C.W. Andersen, M. Marcinek, A. Szczesna-Chrzan, M. Gaberscek, M.R. Palacin, M. Uhrin, J. Friis, Toward a unified description of battery data, *Adv. Energy Mater.* 12 (2022) 2102702, <https://doi.org/10.1002/aenm.202102702>.