CellPress

Joule

Perspective

The challenge and opportunity of battery lifetime prediction from field data

Valentin Sulzer,[1] Peyman Mohtat,[1] Antti Aitio,[2] Suhak Lee,[1] Yen T. Yeh,[3] Frank Steinbacher,[4] Muhammad Umer Khan,[5] Jang Woo Lee,[6] Jason B. Siegel,[1] Anna G. Stefanopoulou,[1] and David A. Howey[2,7,*]

SUMMARY

Accurate battery life prediction is a critical part of the business case for electric vehicles, stationary energy storage, and nascent applications such as electric aircraft. Existing methods are based on relatively small but well-designed lab datasets and controlled test conditions but incorporating field data is crucial to build a complete picture of how cells age in real-world situations. This comes with additional challenges because end-use applications have uncontrolled operating conditions, less accurate sensors, data collection and storage concerns, and infrequent access to validation checks. We explore a range of techniques for estimating lifetime from lab and field data and suggest that combining machine learning approaches with physical models is a promising method, enabling inference of battery life from noisy data, assessment of second-life condition, and extrapolation to future usage conditions. This work highlights the opportunity for insights gained from field data to reduce battery costs and improve designs.

Context & scale

To enable the transition to a clean economy and ensure confidence in energy storage technologies, advances are required in reliability, safety, and extended usage of batteries. While headline-grabbing improvements have been made in battery materials, significant advances may also be achieved in managing behavior via enhanced modeling and real-time sensing. These are often chemistry-agnostic and hence can be coupled with future materials and next-generation chemistries such as lithium metal.

Here, we explore how physics-based and data-driven modeling informed by measurements from end-use devices enables new battery lifetime models. Although challenging, this will lead to reduced costs by reducing the battery size needed to satisfy warranties and guarantee performance. It will also elucidate degradation mechanisms, improving safety and reducing downtime by enabling appropriate interventions. Finally, it will inform decisions on second life, enabling a circular economy for batteries.

INTRODUCTION

Batteries are used in a wide variety of applications, from consumer electronics to electric cars, rail, marine, and grid storage systems. A critical need for consumer acceptance in electric vehicles is to achieve longer range and lower cost via pack size reduction.[1,2] All of these objectives depend on accurate state of health (SOH) estimation and predictions of lifetime under various operating conditions. More accurate lifetime prediction improves battery technology at all stages of a battery's life. First, it can shorten the product development cycle, for example, by elucidating failure mechanisms, in particular, if models can be incorporated in a closed loop with experiments.[3] Second, it can be used to optimize manufacturing protocols. Third, improved lifetime prediction can lead to lower warranty and insurance costs, timely preventative maintenance, lower up-front capital cost by reduced over-engineering, and better control of charging and discharging that could prolong life.[4] Finally, it leads to improved prospects for second-life applications—supporting the creation of a circular economy around battery manufacturing, re-use, and recycling will also be critical as demand is forecast to outpace raw material supply and refinement over the coming decade.[5]

The criteria for determining end of life may vary by application, but generally this occurs when the battery can no longer meet the requirements of range, operating time, or maximum power capability under typical usage profiles. The key parameters that affect end of life are capacity (available energy) and internal resistance (available power).[6] Battery aging depends on intrinsic factors, such as manufacturing variability and pack design, and extrinsic factors, such as temperature and intensity of usage,[7,8] and is therefore difficult to predict, particularly outside of the laboratory.[2] Existing reviews and perspectives in the literature consider methods for SOH prediction,[9–15] lifetime prediction,[9,14,16,17] and/or fusion of physics-based and data-driven models,[17–19] but are historically restricted to fairly small datasets under very controlled conditions. Complementary to existing literature, this perspective examines the unique challenges of battery lifetime prediction with field data and for second-life applications and reviews which approaches are most promising for addressing these challenges. This first requires a review of methods using lab data (life prediction from lab data). We then lay out the challenges and assess promising methods for field data analysis (life prediction from field data), where we discuss the additional value that field data offers for lifetime prognostics and the difficulties in obtaining and processing these data. Finally, we address high-throughput testing for second-life asset evaluation (life prediction for second life).

In a lab setting, which is the best understood and most studied in the literature, the cycling pattern of batteries can be closely controlled and regular reference performance tests (RPTs) can be performed to quantify health. However, field data from real-world applications exhibit irregular cycling patterns, varying operating conditions, and path-dependent degradation mechanisms, making reliable predictions difficult. This setting is extremely relevant for industrial needs, such as prediction of the remaining useful life of a customer's electric vehicle or compliance with warranty conditions for grid storage systems, but prognostics using real-world data remains an open research challenge.

Industry sectors such as automotive manufacturers have narrow profit margins and comprehensive certification requirements that necessitate extensive laboratory testing, therefore gathering fleet data may come at an additional cost and effort that is hard to justify. Since test data are often not available across the wide range of cells used in packs commercially, a basic set of lab degradation measurements is beneficial as a starting point for understanding the impact of operating parameters on degradation. Additionally, given that automotive cells are considered degraded when they reach 80% SOH (i.e., of their initial capacity), there is a need for capacity and resistance estimation accuracy of at least 5% and ideally 2% to underpin lifetime prediction. Defining the accuracy and confidence levels necessary for a health-conscious battery management system is still an urgent research goal.[20]

To build an accurate, general model of battery behavior that covers many usage conditions, a large amount of aggregated data from a population of users is required—it is insufficient only to do this on an individual end user basis. The information that is gathered from intelligently tracking degradation at the fleet level in the field could be used to improve user experience for individual battery end users via over-the-air software updates although there may be regulatory barriers inhibiting this. Finally, at the end of its first life (e.g., in an electric vehicle), a battery may be assessed for possible second-life application (e.g., grid storage). Estimating health at this point comes with additional difficulties, such as potential lack of historical data and a change in the future aging mechanism of the battery due to different operating conditions in second life. Controlled RPTs are possible and can be designed around screening and techno-economic analysis of batteries for second life but are time consuming and require equipment and space that translates to cost, disadvantaging the economics of re-purposed batteries.

To address the challenge of lifetime prediction, three general approaches exist—empirical aging maps, data-driven models, and physics-based models—and later

[1]University of Michigan, Ann Arbor, MI, USA
[2]University of Oxford, Oxford, UK
[3]Voltaiq, Berkeley, CA, USA
[4]Siemens, Erlangen, Germany
[5]Continental AG, Regensburg, Germany
[6]Samsung SDI Company, Yongin-si 17084, South Korea
[7]The Faraday Institution, Harwell, UK
*Correspondence: david.howey@eng.ox.ac.uk
https://doi.org/10.1016/j.joule.2021.06.005

in this paper, we also introduce a fourth technique that combines physics-based and data-driven models. Empirical aging maps model capacity fade explicitly as a function of time or charge throughput, parametrized by operating conditions such as temperature, C-rate, and depth of discharge.[7,21–25] In the purely data-driven approach, very few assumptions are made about the underlying principles governing the behavior of the battery, and machine learning models are trained using raw input signals (current, voltage, and temperature).[26–36] Another type of data-driven method uses preprocessed features from the voltage, current, temperature, impedance, or power curves as inputs to machine learning models.[36–54] Finally, in the physics-based approach, models are constructed from first principles, with tuning parameters found using a relatively small number of experiments. Such models include differential voltage analysis models,[55–60] equivalent circuit models (ECMs),[7,61–63] and first-principles degradation models based on porous-electrode theory.[8,64–70]

Unfortunately, neither the data-driven nor physics-based methods alone can solve the challenge of battery lifetime prognostics from field data. The challenges facing physics-based modeling have been well documented, such as a large number of coupled and nonlinear degradation mechanisms,[8] which evolve nearly unobservably from the electrical measurements and are difficult to parametrize. As a result, not all researchers agree on which degradation mechanisms to model and how to implement them, since many different mechanisms and formulations can explain similarly observed degradation behavior.[8] We note that mechanical[71] or acoustic[72] measurements have been proposed as a means to address the lack of observability, but such measurements are not yet widely used.

Meanwhile, data-driven approaches suffer from the "curse of dimensionality," where the amount of data needed to capture all combinations of operating conditions grows quickly with the number of conditions being investigated. This is compounded by the relatively slow rate at which battery lifetime data can be acquired, taking several months or years of experiments for each change in chemistry (e.g., electrolyte additive), form-factor, or manufacturing process. Furthermore, trade secrecy limits how much data are available to individual researchers. The inherently nonlinear, path-dependent nature of battery degradation further exacerbates this problem.[59,73–76]

Here, we start by introducing different approaches for life prediction in lab settings (life prediction from lab data). The advantages and disadvantages of each method in terms of computational complexity, data requirements, and accuracy are also discussed. Then, the challenges and opportunities of applying life prediction methods to field data are presented (life prediction from field data), and we suggest that hybrid methods that combine physics-based and data-driven approaches show promise for this because they combine accuracy, robustness to limited or low-quality data, and generality (life prediction from field data). Finally, in life prediction for second life, we discuss the challenges of lifetime estimation for second-life applications. These include assessing viability of batteries without having historical data available, in particular determining whether or not the battery has degraded beyond the knee point (i.e., the point beyond which degradation may accelerate toward end of life) and understanding the cost-benefit between testing re-purposed batteries versus the increased revenue from more accurately knowing their SOH.

LIFE PREDICTION FROM LAB DATA

In a laboratory, battery cycling can be repeated consistently, and conditions such as current and temperature can be closely controlled (Figure 1). Measurements of

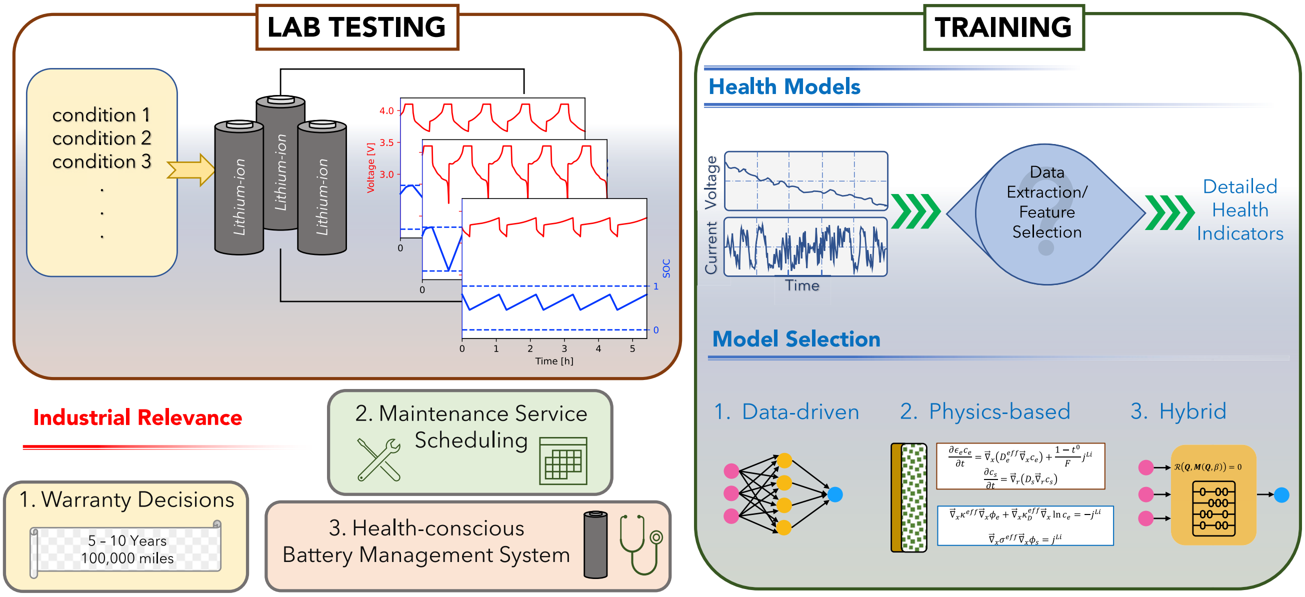

Figure 1. Prognostics using lab data

Summary of approaches for prognostics using lab data. A variety of use cases can be tested and health models trained using data-driven, physics-based, or hybrid paradigms. These models are typically developed to underpin warranties but are not validated beyond the warranty period or in other usage scenarios.

"ground truth" battery health—capacity and resistance—can easily be taken with regular RPTs as required, and batteries can be cycled continuously until end of life, which is often specified as the point where measured capacity reaches 80% of the pristine cell capacity.[77]

This means that laboratory tests are useful to build models that explore how different operating conditions affect cycle life, and hence determine warranties, maintenance schedules, and system sizing. The operating conditions used in laboratory tests include driving patterns and schedules tailored to very specific applications, such as a specific electric vehicle model. Laboratory tests can also be used to guide the development of new battery chemistries, optimize battery design, and improve manufacturing processes, for example, by investigating the effect of parameters such as humidity or formation cycles on performance.

Battery degradation testing is a lengthy process, therefore extreme operating conditions, such as high C-rates or elevated temperatures, are often used to accelerate aging. Even with accelerated aging, it can be slow to assess the degradation impact of individual manufacturing parameters such as materials and processing choices, design factors such as cell size, number of layers and electrode thicknesses, and formation protocols. Additionally, since different manufacturing parameters interact non-linearly, varying each parameter individually may not tell the full story. Therefore domain-knowledge is required to ensure that testing efforts are as effective as possible.

Significant research has been undertaken on modeling lab battery test data, and it can broadly be divided into four categories, summarized in Figure 2: empirical aging models, pure data-driven methods, feature-based data-driven methods, and physics-based methods. In this section, we review state-of-the-art approaches within each of these categories.

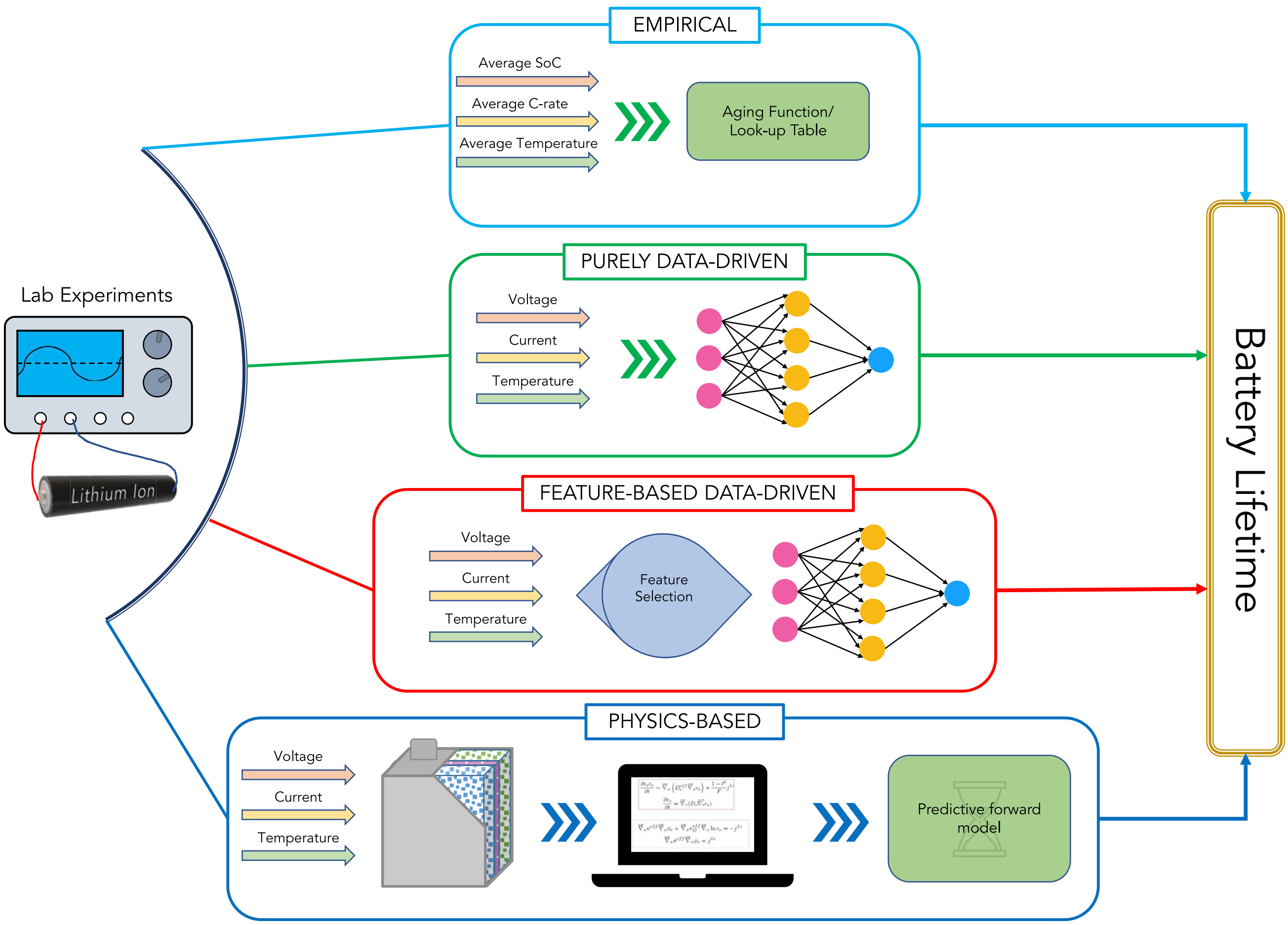

Figure 2. Different approaches to battery lifetime estimation
Empirical, purely data-driven, feature-based data-driven, and physics-based approaches to battery lifetime prediction.

Empirical aging models

A typical measure of battery degradation is the capacity fade curve, which describes how the capacity changes as a function of charge throughput, equivalent cycle number, or time. Therefore, the simplest approach for lifetime prediction is to build an empirical model of the capacity fade parametrized by operating conditions. This may depend on a number of factors including time and charge throughput. As first proposed by Bloom et al.,[21] such empirical models usually have a square-root-of-time dependence, due to diffusion-limited solid electrolyte interphase (SEI) formation, and Arrhenius kinetics for the temperature dependence.[7,21–25] Further refinements include accounting for C-rate,[22–24] average state of charge (SOC),[23,24] depth-of-discharge range,[7,24] and voltage.[7,25] Furthermore, recent models separate capacity fade into calendar aging, which depends on time, and cycle aging, which depends on charge throughput.[7,23,24]

This approach is simple and easy to implement, making it common in industry for developing maps of lifetime from lab data. However, it has a few limitations. First, very large amounts of data are required to interpolate over all operating conditions. Second, this approach cannot easily account for cell-to-cell variations due to manufacturing or heterogeneous current and temperature distributions within a pack. Third, separate maps are usually developed for cycle and calendar aging, which misses the interactions between these,[73] whereas the methods that we will review below can incorporate these interactions as long as appropriate data are included in the training sets. Finally, empirical aging maps may fail to capture knee points in the capacity fade curve.[7,24] Knee points are a change in the degradation rate likely caused by a change in the underlying mechanism[78] (e.g., SEI growth later leading to lithium plating[66]), and they are a particular challenge for prediction, especially with simple empirical models.

To overcome these limitations, prognostics can be improved by also taking into account the individual raw time series data (such as voltage) of the cells during cycling, instead of just the capacity obtained during the RPTs.

Purely data-driven models

This approach consists of using measurements such as the current and voltage directly as inputs to a machine learning model in order to learn the remaining useful life as the output. While several studies have shown promise in estimating the present SOH of the battery,[26–32,35,79] there is limited application to date of these methods to prognostics, with one example being Zhang et al.,[33] who use raw electrochemical impedance spectroscopy data to predict remaining useful life (RUL). This may be due to the lack of available data: while each cell has hundreds of cycles to use as training data for SOH estimation, its cycle life is just a single data point per cell.

Another possible approach is to train a machine learning model for "forward simulation," where the model learns how much capacity fade occurs during short intervals based on the existing capacity, current, or temperature during the interval.[34–36] Then the RUL can be predicted by adding together the capacity fade from all the intervals under typical usage conditions and seeing where the resulting trajectory crosses 80%.

Feature-based data-driven models

In this approach, there is a preprocessing step in which features are extracted from the voltage and current, informed by physical understanding of the cell's behavior. These features are then used as the inputs to machine learning algorithms. Similarly to pure data-driven methods, much of the literature focuses on present SOH estimation rather than prediction of future SOH.[36–44,46] However, since the models that take features as inputs are less complex than those that take raw data as input, they require less training data than pure data-driven approaches and are more readily applicable to lifetime estimation. For example, Chinomona et al.[47] use various statistical features from the voltage, current, and temperature, while Yun et al.[48] and Greenbank and Howey[49] use as features the time spent within certain voltage, current, power, and temperature ranges.

While those studies use features from the current or voltage within a single cycle, Severson et al.[52] showed that using features generated from changes between different cycles can give very accurate predictions of lifetime, even using a simple regularized linear regression model. Fermín-Cueto et al.[53] improved the accuracy of this method (lower prediction error with fewer cycles) by using more features and a more advanced machine learning algorithm. In a previous paper,[54] we applied this method to a different dataset for NMC/graphite cells, achieving an accurate prediction of battery lifetime, and showed that these features are closely correlated with loss of lithium inventory in the cells, demonstrating the importance of a physical basis for the features.

Accurate data-driven prediction of the knee point in the capacity fade curve using feature-based data-driven models has recently been demonstrated,[49,53,80] despite the fact that data-driven methods may fail to extrapolate correctly under changing usage conditions due to path-dependence[59] and may not even correlate with aging unless a certain depth of discharge is reached during usage.[54]

Finally, some feature-based "forward simulation" approaches have also been proposed, using physics-informed features as inputs to Gaussian processes to predict incremental capacity fade and, hence, RUL under typical operating conditions.[45,50,51]

Physics-based models

This broad category includes atomistic models, continuum approaches based on porous-electrode theory, through to ECMs. We include ECMs here as "lumped" physics models using electrical engineering components to capture the electrochemical behavior.[63,81] In terms of degradation, circuit models can be used to estimate and track parameters empirically as cells age.[7] Examples of lifetime prediction using circuit models are Zhang et al.[61] and Chu et al.[62] who employ observers to identify the internal states of an ECM and then fit an empirical aging map to find future changes in capacity. A challenge with these approaches is that careful tuning is required to achieve robust performance. To address this, Aitio and Howey[82] show that applying Gaussian process regression to identify functional dependencies of model parameters can give smoother and more dependable results when using drive-cycle data. ECMs are popular since they are easy to implement and parametrize, although recent work has shown that similar computational efficiency can be achieved with models based on porous-electrode theory.[83–85]

The other relevant category of physics-based models is those derived from first principles—typically continuum approaches using porous-electrode theory. These may be extended to capture the underlying degradation mechanisms that cause capacity fade, such as SEI layer growth,[8,65–70,86–88] lithium plating,[8,65–67,89,90] and particle swelling and cracking.[8,91] These models can be used to directly simulate the entire life of the cell under certain operating conditions.[8,64–70] In theory, such models could be used for prognostics by parametrizing them with early-life data for a particular cell and then simulating the remaining life of the cell and seeing where 80% capacity is reached. An advantage of this approach is that it also provides the predicted cause of failure, enabling remediation strategies before end of life such as tightening of safety limits in the battery management system.

The main weakness of these models is their lack of flexibility and parametrization difficulty. While excellent agreement with the available data can be achieved, it is not possible for a mechanistic model to account for every single eventuality. For example, even when a large number of degradation sub-models are included, some aspects of the experimental dataset still cannot be fitted accurately.[8]

In the following section, we suggest ways in which physics-based models can be made more flexible (and hence model a wider range of degradation of mechanisms) by augmenting them with methods from machine learning.

Accurate parametrization of continuum physics-based models for a fresh cell is challenging, requiring cell teardown and specialized testing equipment[92,93] to determine a wide range of parameters such as reaction rate constants, conductivities, diffusivities, particle sizes, etc. Determining the degradation parameters, such as SEI kinetic parameters, by inverse modeling is even more challenging due to the computational time required to simulate the entire lifetime (e.g., 15 min with Ouyang et al.'s relatively simple model,[64] compared with microseconds for an empirical aging map). For physics-based models to become more widely adopted for lifetime estimation, this computational time needs to be reduced, either through the development of reduced-order models[83] or numerical methods. Advancing efficient physics-based models is one of the goals of open-source battery modeling framework PyBaMM.[94]

Physical understanding can also be used to define more specific degradation metrics than capacity or resistance. For example "degradation modes," such as loss of active material and loss of lithium inventory, may be estimated from data or models and linked to different underlying mechanisms, such as SEI growth or lithium plating.[60] These can be identified using differential voltage analysis[55–59] and used for prediction. For example, Hui et al.[95] use differential voltage analysis to identify degradation modes, then predict future change in these degradation modes—achieving better accuracy than empirical aging maps.

LIFE PREDICTION FROM FIELD DATA

Lab battery testing is limited in the number of test channels available and the time available for tests. Ultimately, what matters is battery performance in real applications. If field data from batteries in end-use applications could supplement lab performance and lifetime tests, this would significantly increase the amount of data available, accelerating our understanding and closing the gap between lab and end-use. It would also ensure that lifetime prediction algorithms are relevant to industry applications. Figure 3 summarizes the aims and impacts of lifetime prediction from field data.

Field data versus representative drive-cycle testing in the lab

One existing approach to bridge the gap between standard lab tests and field data is to test batteries using representative loading patterns in the lab. To this end, many researchers have characterized "typical" user driving patterns.[2,96,97] The advantage of this approach is that it can be performed in the lab with high-accuracy equipment, controlled conditions, and frequent characterizations. However, complementing this with field data is still very valuable for a number of reasons. First, there are never enough testing channels in the lab, and they are never available for long enough to test all combinations of conditions necessary—whose number grows factorially with the number of aging parameters. In contrast, field data may be available at relatively low cost since the cells are already being cycled, and (by definition) the entire range of operating conditions is covered.

Second, lab testing is contrived and different to real-world usage. Due to constraints of both time and testing equipment availability, lab tests usually impose extreme conditions and short rest periods between cycles. This can lead to under-prediction of battery lifetime, and hence over-engineering of pack design. Accelerated aging experiments often stress the battery more intensively than standard cycling in the field. This is accounted for by experimentally developed maps between accelerated experiments and real usage scenarios. However, these maps take a long time to obtain, because they require replicating real usage scenarios in the lab, and even then it is difficult to perfectly relate lab tests to field data. For example, it has been shown that rest periods between cycles can be beneficial for the overall life of the battery,[73] but lab experiments cannot always incorporate long resting periods due to testing time constraints. For extension of model validity from lab accelerated aging tests to real-world usage, it is important to standardize methods to interpret

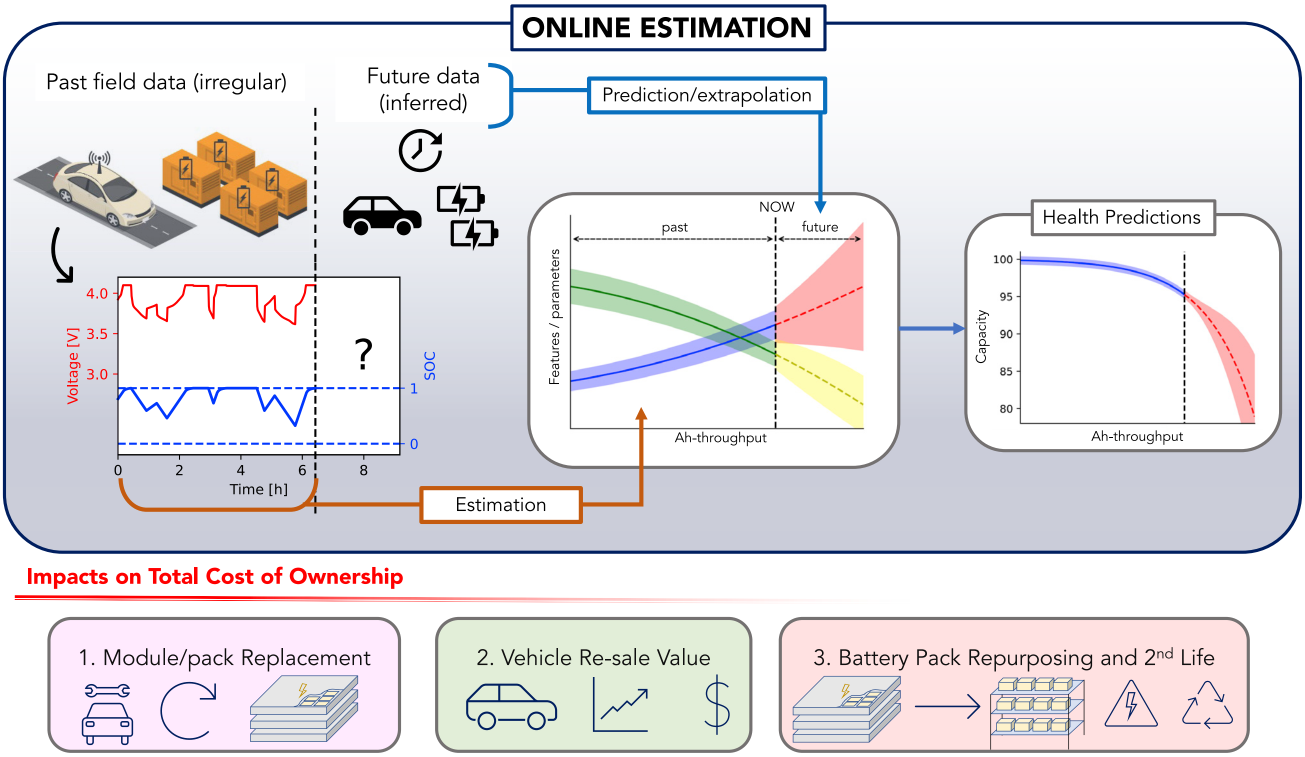

Figure 3. Prognostics using field data

Summary of the aims, processes, and impacts of state of health prediction from field data. Challenges include noisy and missing data, uncontrolled partial cycling conditions, lack of regular validation tests, and uncertainty over future usage.

not just accelerated cycle aging data but also accelerated calendar aging and mixed cycle plus calendar aging scenarios. For both lab-accelerated-aging and field-data-aging research, it is important to rigorously understand the impact of cycles-per-day and rest-time conditions on lifetime.

Third, there could be exogenous factors in the field that are not accounted for in the lab, such as seasonal temperature variations or mechanical vibrations causing failure. Fourth, from the point of view of understanding battery lifetime, which depends on both intrinsic (manufacturing variability) and extrinsic (usage) factors, more data are always valuable because it increases the statistical confidence with which we can build lifetime and performance models.

Challenges and opportunities of field data

Three main issues arise in this context: the wide variety of uncontrolled usage scenarios, data quality issues, and lack of clear validation data. First, unlike lab-based prognostics where operating conditions are held constant or at least controllable, in the real world, operating conditions can vary throughout a cell's life. Examples for electric vehicle batteries include temperature differences due to seasonality, cell-to-cell variations within a battery pack, or varying SOC windows associated with trip distances. Depending on driving conditions, charging and discharging current/power levels can affect battery degradation to a greater or lesser extent. Recent studies have shown that battery degradation may also depend on the order of the operating conditions, a phenomenon known as "path- dependence" in the literature.[59,73–76] Another source of uncertainty in lifetime predictions arises because future operating conditions themselves are uncertain.

Table 1. Summary of existing state of health estimation papers using real-world pack-level battery data

Ref	Dataset Used	Aim	Method
Song et al.[99]	700 B/HEVs, Shanghai	SOH estimation	neural network
Wang et al.[100]	8,032 EVs, Beijing	evolution of cell variability	ECMs and regression
She et al.[101]	18 EVs, Foshan	SOH estimation	ICA and neural network
Huo et al.,[102]	16 EVs, Beijing	SOH estimation	empirical model & Bayesian network

Second, field datasets are difficult to work with, for several reasons. Cell-to-cell variations within a pack or modules, less accurate sensors, noise, extreme C-rates, smaller coverage of the available range of conditions (e.g. microcycling of SOC), missing or false data, and rapidly varying currents make it very challenging to extract features or fit parameters.[54,98] There may also be long periods of missing data, for example, temperature data while the battery is at rest (and hence not recording). These factors also restrict the identifiability of both mechanistic and data-driven models. For physics-based models, the input signal is not rich enough for parametrization and, therefore, simplifications of the model are required. On the other hand, performance of purely data-driven methods is limited by the extent that the training data covers the possible space of inputs, since extrapolation outside of the observed data comes with significant uncertainty. The third difficulty for real-world prognostics is the challenge of verification or validation, i.e., in field data there is typically a lack of ground truth against which to compare predictions. This is due to the lack of controlled RPTs, which makes capacity and end of life difficult to validate, and the fact that not all cells are cycled to end of life, reducing dramatically the data fidelity and availability. A related issue is how often to retrain a prognostic model, where a balance needs to be struck between having sufficiently rich data versus capturing gradual changes in usage and operating conditions.

In the remainder of this section, we review existing studies on this topic, then suggest approaches that can be used to overcome the issues and address the challenge of prognostics in the field. We separate these again into "data-driven" and "physics-based" methods, though most of these methods are hybrid in nature. The challenges in this section represent an exciting new frontier for battery system development and demonstrates real-world application of hybrid data-driven techniques. We finally discuss two practical aspects of implementation, namely data management and uncertainty.

Existing studies using field data

In comparison to studies based on lab generated data, existing literature using field data for life modeling is sparse and is mainly focused on diagnostics rather than lifetime prediction. However, some large databases of electric vehicle usage data have been collected and analyzed, as summarized in Table 1. These existing works focus on data-driven methods to estimate SOH. For example, Song et al.[99] used operating data from 700 electric vehicles to fit a feed-forward neural network using features calculated from usage history as inputs to estimate capacity, where the target output itself was estimated from partial charging curves. Alternatively, Wang et al.[100] used a larger dataset from 8,032 electric taxis to track cell-to-cell variability at the pack level over time. The same data source was used by She et al.[101] to estimate capacity loss in 18 electric city buses by fitting a neural network with inputs consisting of usage features and the output of the peak value in the incremental capacity curve during charging. A different approach was proposed by Huo et al.,[102] who installed data collectors on 16 electric taxis and performed periodic capacity calibration tests to obtain validation data for probabilistic SOH estimates obtained through a Bayesian network with an empirical degradation model. These works give a glimpse of what is possible using large datasets, and they also encounter the challenges mentioned in this perspective, such as missing or noisy data and lack of validation. These papers do not make their data and code available for others to download, so it is not possible to reproduce their results. The lack of availability of large publicly accessible datasets, including validation data and real-world operating conditions, remains a severe limitation for model development and performance quantification. Future studies on field data should strive to make their data and methods openly available[103] where possible to enable comprehensive quantitative comparison of algorithms.

Feature-based data-driven approaches

Data-driven approaches informed by physics show promise for use in field applications, particularly methods that have been demonstrated to work with relatively short bursts of data, data from controlled charging, and feature selection approaches that can handle varying operating conditions.

For SOH estimation, the Gaussian process method proposed by Richardson et al.[44] gives good results with only 10s of constant-current data. Similarly, methods that rely on features from partial SOC windows[47–49] are more likely to work with field data since conditions are more variable. Approaches that are based on extracting features from the battery charging phase (either the constant-current[37,42] or constant-voltage phase[37–39,42]) are of interest for field applications, since in some cases, such as electric vehicles, the charging phase can usually be closely controlled while the discharge phase depends on the power demands of the user. However, these methods still require the charge to begin from a small enough SOC for the features to exist, which may not be the case when microcycling (for example, regular short commutes in an electric vehicle). In this case internal resistance may be a more reliable health indicator than capacity.

For prognostics, methods based on incremental updates of capacity fade[35,36,45,50,51] may work well with field data since they are able to handle rapidly varying operating conditions, usually using some form of averaging.

To be effective with field data, feature-based data-driven methods should demonstrate that extracted features are independent of instantaneous operating conditions such as C-rate, temperature, or SOC. If this is not the case, then the prognostics algorithm must account for such variations, which requires large amounts of training data. For example, if the map from "time-between-voltages" to RUL depends on the C-rate, then enough tests need to be run to capture this dependency and the number of tests required grows exponentially with the number of such dependencies. In contrast, physics-based models directly encode these dependencies into the model. We now explore how such models can be used for life prediction in the field.

Hybrid physics-based/data-driven models

As mentioned previously, physics-based degradation modeling aims to simulate the physical mechanisms that cause the battery to age. In an ideal world, we could perfectly model each of these mechanisms and, hence, construct a digital twin of a battery which would tell us exactly how it will age in future.

In practice, most physics-based degradation models that have been proposed are only validated against data from a single operating condition, and studies that attempt validation using several operating conditions indicate that it is difficult to generalize.[8] This suggests that degradation models could be enhanced with data-driven methods in order to accurately predict a wide range of possible aging mechanisms and cell-to-cell variability.[104] To this end, promising frameworks exist, including the field inversion and machine learning paradigm,[105] physics-informed neural networks,[106–108] neural ordinary differential equations,[109] universal differential equations,[110] and neural differential equations.[111] A key advantage of these methods, as demonstrated for modeling the spread of COVID-19 by Dandekar et al.,[112] is that they extrapolate beyond the training data much more accurately than pure data-driven methods. This suggests that these methods could be particularly well suited for battery lifetime estimation. Such hybrid methods have been applied to battery models by Bills et al.,[113] who add unknown degradation terms to the model and then learn these terms using neural networks, and by Tait et al.[114] who use latent forcing functions to improve the accuracy of a simple electrochemical model. For a further in-depth review of these hybrid physics-based/machine- learning frameworks, we point the reader to Aykol et al.[17]

Hybrid models outside of these frameworks are also possible. For example, Zhang et al.[61] use an equivalent circuit model to infer capacity from non-constant-current data, then use this to parametrize a capacity fade curve. Aitio and Howey[82] use a Gaussian process to accurately infer the SOC dependency of resistance from synthetic drive-cycle data.

These studies demonstrate promising early results, with further research needed in this direction. A key issue, however, is lack of openly available field data. While hybrid methods will be most useful when applied to field data, they should also be demonstrated using lab data where results can be more easily validated.

Data management

We have demonstrated that collecting and analyzing data from the field can be beneficial but doing so requires investment. In our view, the costs and benefits of extensive data collection will vary depending on the application and stakeholder. For safety, lithium-ion systems already have extensive voltage, current, and temperature monitoring; recording this data and/or uploading it to a server comes at an additional cost, and while small relative to the battery cost, this may still be prohibitive depending on the application business case. However, for those who choose to do this, there could be significant benefits in future (e.g., cell supplier selection, lifetime modeling, improved designs, and second-life use). In many applications, data are already collected and stored on a server, especially for warranty compliance of grid storage systems that have a lifetime of 10–25 years and may represent substantial ($10M's) investments. There are indications from industry that more detailed field data collection is very valuable, especially for cell manufacturers. For example, Northvolt has a desire to collect extensive battery data from manufacturing through to end of life,[115] including cell-level data from batteries deployed in real products. This is a lofty goal, but detailed cell level field data are obviously viewed as valuable for improving cell design and manufacturing. Off-grid energy suppliers are also interested in detailed data collection because their systems are often in rural areas that are difficult to reach, and therefore logistics and preventative maintenance (e.g., battery replacement) planning requires detailed remote monitoring and battery health prediction. Even if the benefit of collecting field data is not immediately clear (due to lack of existing methods to analyze it), historical data are invaluable for the validation of new algorithms.

Managing large volumes of data is an important aspect of dealing with battery field data as the number of cells grows from tens to millions. Significant computational

Table 2. Size of data to be uploaded to the cloud

Method	Data uploaded to cloud	Data per cycle per cell
Empirical aging map	aggregated C-rate, temperature, SOC window	bytes
Pure data-driven	time series data	kilobytes
Feature-based data-driven	features	bytes
Physics-based (offline processing)	time series data	kilobytes
Physics-based (online processing)	degradation parameters	bytes

Some methods require high computing power and can only be performed in the cloud, so all of the time series data need to be uploaded. For other methods, some aggregation can be performed by the battery management system, and only the aggregated data uploaded. The type of aggregation depends on the method.

resources are required to store, process, and retrieve data: over the course of its lifetime, a single cell may produce on the order of 100 MB (assuming a thousand cycles of 1C/1C cycling while recording time, current, voltage, and temperature at a frequency of 1 Hz, and 4 bytes per single-precision data point) of time series data. This processing must be performed automatically and requires close collaboration between battery engineers and software engineers. Furthermore, battery data can come from a wide variety of sources: design, manufacturing, lab testing, field use, and second-life use. To maximize the information that can be gleaned from the data, it is important to integrate all the data from this battery supply and operation chain onto the same platform. In particular, manufacturing metadata including materials and processing information is of equal importance to time series performance data and can be invaluable when trying to understand variations between different cells and modules. It may be that not every single product needs to be monitored in this way, just a sub-sample from which population behavior can be inferred.

Once the raw data has been recorded, these can be processed and aggregated in various ways by the battery management system itself before uploading to the cloud/server. Typically, time series data are recorded for each group of cells in parallel (a "brick"), and a module consists of several bricks in series with several modules together making a pack. Data can then be aggregated between these groups. There are two levels of aggregation: compression of time series data and aggregation between bricks. For the first type, depending on the algorithm, it may not be necessary to upload all of the raw data to the cloud/server, and for each cycle kilobytes of time series data may be reduced to just a few bytes of aggregated data (Table 2), for example, by extracting key features. Feature calculation is usually cheap and can be performed by the battery management system, then only the features need be uploaded to the cloud/server. Alternatively, if physics-based (or hybrid) models are simple enough to be run and fitted online, then only key internal aging parameters need to be recorded at each cycle. In other cases, all the raw data must be uploaded to the cloud/server for processing, but once this has been done it can be archived or, in some cases, discarded.

Aggregation between cells can take several forms. The simplest approach would be to record data at the module level (i.e., a number of cell bricks in series) instead of the brick (i.e., parallel cell block) level, which does not take into account any heterogeneity in the module. Fast sampling at the module level or even polling[116] cell level data for individual cell level SOC and SOH estimation requires fast convergence that can only be achieved with the right excitation, potentially during charging.[117] At the other extreme, the data from every brick can be uploaded to the cloud, but

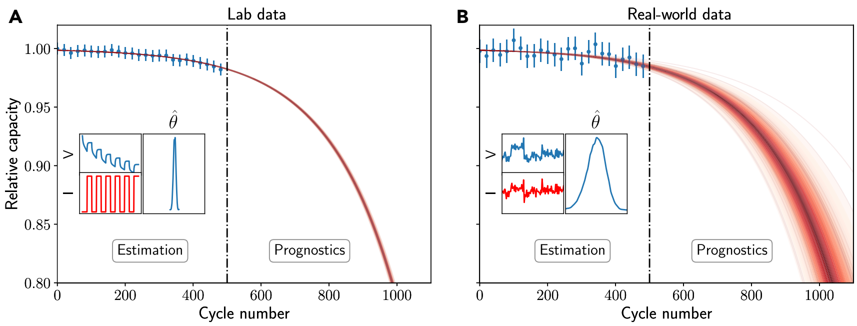

Figure 4. Quantifying uncertainty

A key challenge in life prediction using field data is managing uncertainty. Controlled conditions and precise measurements in lab conditions cannot be replicated in real-world operating environments. This results in less accurate and lower precision estimates of model parameters, $\hat{\theta}$, which have to be accounted for at the prognostic stage. Plots generated with PyMC3.[121]

this may introduce a lot of redundant data. A good compromise, already done by some manufacturers, is to store only max/min/average brick voltages, as well as module level data, in order to capture the heterogeneity in the module. This requires measuring the data for every brick, which is usually already done for safety reasons. However, storing and transmitting this information to a central server may require firmware changes at the local level and may require hardware changes too (depending on how the battery management system is configured). Although health estimation and lifetime prediction algorithms may be applied at either the cell, module, or pack level, accuracy will likely be lower if only system level data are available.[118]

Finally, data processing and storage should take into account privacy issues—for example, an individual's daily routine could potentially be inferred from the power usage of their electric vehicle. Methods to anonymize and aggregate data without losing insights for lifetime prediction should be developed in order to address these issues.

Dealing with uncertainty

Regardless of the chosen method, a key requirement in model fitting and prediction with field data is the rigorous propagation of uncertainty through the estimation algorithms. Due to the constraints imposed by the data, the variance of model parameters in real-world scenarios is inherently large. Illustrated in Figure 4, the uncertainty of health estimates is carried forward to lifetime predictions, where the compounding of uncertainty over future behavior is complex due to nonlinear dependencies on model parameters, path-dependence, and unknown operating conditions. To this end, Bayesian methods for parameter estimation and prediction show promise because they enable principled inclusion of model and observation uncertainty. With these approaches, full parameter posterior probability distributions obtained either by analytical approximation[119] or using Monte Carlo techniques[120] may be used to construct battery performance estimates with realistic confidence bounds.

LIFE PREDICTION FOR SECOND LIFE

When a battery can no longer meet the capacity or power requirements for its primary application, it can sometimes be re-purposed into a different "second-life" application. This is environmentally beneficial and economically viable, especially with the growing market for lithium-ion batteries for electric vehicles, which can then later be re-purposed for grid or stationary storage applications.[122]

The challenge here is to predict the RUL of the battery under its intended new application, which is likely very different to the previous use, in order to decide whether such new usage is viable and inform warranty decisions. There are two particular issues: first, historical data to audit first life usage may be unavailable, for example, due to commercial secrecy, and if available, it may not be of the required reliability or fidelity—although initiatives such as the Battery Passport[123] are being implemented to promote reliable sharing of data. In some cases, time series data from first life and manufacturing metadata are sold with the battery at a premium. Second, operating conditions can vary significantly from first life to second life, for example, grid applications will typically have different thermal management systems and reduced SOC windows and C-rates compared with EVs, so a model of performance from first life may not be valid for second life. However, it is possible to perform controlled cycles at the re-purposing stage, such as slow deep discharges to determine capacity, pulse tests to determine internal resistance, or electrochemical impedance spectroscopy to determine the internal parameters of a model. The RUL estimation must account for the fact that the usage patterns in the second life will be different (usually, less intensive) than they were in the first life. Therefore, even if historical usage information is available, prognostics algorithms must be more advanced than simply extrapolating aging patterns and models from the first life.

Assessing state of health for second life

At its simplest, assessing the second-life viability of a particular battery pack consists of measuring its SOH, for example, capacity under constant-current discharge at various C-rates. However, this may not tell the full story about the second-life viability of the pack, due to the intrinsic variability and path-dependence of capacity fade.[73] Cells that have the same SOH at a point in time may degrade at different rates in the future even under the same operating conditions. In particular, Martinez-Laserna et al.[124] show that it is crucial to determine whether or not the battery has passed the knee point in the capacity fade curve, since degradation does not slow down after this, even under less intensive usage conditions. If historical data are unavailable, it is not at the moment clear how to determine whether the cell is past the knee point. This is an interesting challenge problem for the community.

To address this, advanced RPTs may be possible to estimate parameters of a physics-based model at a level of detail that goes beyond just estimating capacity and resistance. It may be the case that some parameters of a physics-based model correlate well with whether or not the battery is past the knee point. Finding such parameters, and ways to easily identify them, would be extremely useful for assessing the second-life viability of individual batteries. In this context, for porous-electrode models, Park et al.[125] suggest a framework for devising current excitation profiles that maximizes the observability of the underlying model parameters, using a sensitivity analysis and Fisher information matrix. Alternatively, electrochemical impedance spectroscopy may be used to estimate parameters of a physics-based model,[126] although without a reference electrode there may be ambiguity in attributing parameters to one or the other electrode. Differential voltage analysis is also a particularly promising framework here, since performing deep low C-rate charges or discharges is possible, and the degradation modes that are identified by differential voltage analysis have been shown to correlate well with knee points.[59,127] A challenge with all of these techniques is that they are usually undertaken on individual cells—implementation at pack level is an open issue.

Cost-benefit analysis for second-life testing

When considering what tests should be performed to assess second-life viability, there is a cost versus accuracy trade-off, and the level of accuracy required for second-life assessment depends on the intended usage. A very simple assessment could consist of a binary go/no-go classification of whether a battery will be usable for the desired second-life application and duration, in which case relatively low-test accuracy is sufficient. A more complex test could also identify the operating conditions that will maximize the battery lifetime in second life, such as SOC window and C-rate, but such tests will require higher accuracy and longer duration. If pack-to-module balancing is planned, very accurate tests are required in order to identify the strongest and weakest packs.[128] In each case, the cost of testing should be no greater than the additional revenue generated by improved test accuracy. While the cost of testing is relatively easy to quantify, calculating the revenue resulting from improved control is an open research question.[4]

Finally, an important question to consider in second-life testing is cell-to-cell variability within a pack. It has been shown that the labor involved in opening up a pack is more expensive than the benefit from testing and screening the individual cells.[129] Therefore, second-life testing strategies should aim to predict the life and identify the performance variability of the cells within a pack without having to test each cell independently. Since most packs measure voltage for every cell brick, this is in principle possible, but would require every single cell voltage to be recorded in a database—at present this is not necessarily the case.

SUMMARY AND OUTLOOK

Battery costs and customer confidence must be improved if we are to improve renewable energy integration by expanding grid storage and rapidly replace combustion engine vehicles with electric vehicles as part of the solution to maintain climate warming within 2°C of pre-industrial levels. To address this, there is a significant need to estimate battery health accurately, diagnose degradation, predict lifetime in different usage scenarios, and detect failures. Combining data-driven and physics-based models will outperform pure data-driven or pure physics-based approaches by allowing robust extrapolation of future behavior from available data.

Different approaches are suited to different use cases. For warranty design and certification of performance, consistent lab test data are used to inform empirical models for life prediction and may also be used for data-driven or physics-based modeling. On the other hand, in most real-world cases where data are less regular, the development of new hybrid models that combine the strengths of physics-based and data-driven methods is essential. Such models are claimed to require less training data than purely data-driven methods, have been shown to extrapolate more accurately,[112] and are more flexible and generalizable than purely physics-based methods. Finally, ensembles of different models could also be considered,[130] since they can outperform the best individual models.

Fleet data can also be used by companies to adjust their models developed with lab data. Developing and training predictive models for battery lifetime requires large amounts of data covering the whole range of operating conditions. Due to time and equipment constraints and the wide variety of operating conditions, lab data cannot completely cover the required range with realistic (non-accelerated) operating conditions. We believe that fleet data from devices in use augments lab data, leading to use-case-specific battery lifetime prediction models and improved designs.

Additionally, deeper characterization can be performed on cells that have reached the end of their life in order to determine their true SOH and cycle life, and hence build up a database for supervised learning. Individual cells that exhibit particularly interesting behavior can be selected for teardown experiments to reveal how these cells aged and hence verify predictions from physics-based models. This can then be used to improve the models for cells that are still in operation. Building and maintaining large databases of field data is a challenging undertaking but is a worthwhile endeavor to prepare for a time when algorithms to analyze it will be available and will require training data. For example, a field data collection effort could begin with collecting data from taxis and buses or vehicles used by company employees. Even data obtained during charging could be extremely useful to update lab-based models. This is already being done in some cities, as noted in life prediction from field data. Open sharing of such data and analysis software between researchers would greatly accelerate efforts to improve algorithms for lifetime modeling. However, collection and curation of this type of data is challenging due to trade secrets and industry competition concerns. A large cross-institutional research project involving national laboratories, universities, and others could facilitate widespread data collection and appropriate anonymization. Companies may also be willing to release small subsets of data openly, for example, through competitions to assist in recruitment and marketing.

To enable these exciting opportunities, further research in several areas is required, as well as enhanced collaboration between disciplines.[131] We highlight the following items as particularly necessary:

(1) Development of consistent, large-scale open-source databases for battery performance and lifetime data and to train and validate algorithms for lifetime prediction. Ideally, these should encompass lab data and field data and share common standards for data and metadata.[103]
(2) Fast and accurate physics-based models, with a particular focus on understanding what are the key degradation sub-models that are important, and how to estimate and track key parameters. Traditionally, such complex models have required very high levels of expertise to implement, but this is becoming more accessible through open-source modeling packages such as PyBaMM.[94]
(3) Scalable algorithms for combining data-driven methods with physics-based models. Some promising pathways for this are the field inversion and machine learning paradigm,[105] neural differential equations,[111] Gaussian process state space models,[132] and universal differential equations.[110]
(4) Methods to diagnose whether a cell has aged beyond the knee point, without requiring historical data, to assess whether the cell can be used in second-life applications.
(5) Quantification of value (e.g., additional revenue) resulting from improved SOH estimation and life prediction, for example, for second life.
(6) Solutions for data privacy issues—anonymization and intelligent aggregation of individual user data without loss of insights for diagnostics and prognostics.

There are many areas of battery research where modeling and data-driven techniques can add value, including materials discovery, battery design, and fast charging algorithms.[133] Achieving breakthroughs in battery lifetime prediction in real applications will require new experimental approaches for lab tests that massively reduce the time required to understand degradation processes in batteries. It will also require new methods to map from accelerated aging tests to real-world conditions—informed by a large amount of field data together with scalable modeling approaches that also account for uncertainty in sensors and models. There is a significant opportunity through this to reduce costs by reducing pack sizing and stretching the usage envelope, in order to accelerate the roll-out of batteries, underpinning the rapid transition to clean energy and transport.

ACKNOWLEDGMENTS

We are grateful for helpful conversations with Dr. Manfred Baldauf (Siemens), Mr. Igor Neiva Camargo (Continental Automotive), and Dr. Markus Schweizer-Berberich (Vitesco Technologies) about the topics in this paper. D.A.H. and A.A. acknowledge funding from the Faraday Institution (EP/S003053/1, grant number FIRG003) and Shell Foundation (agreement 22077). A.G.S., V.S., and P.M. acknowledge support from the University of Michigan Battery Lab. This work was partially supported by the National Science Foundation under grant no. 1762247.

AUTHOR CONTRIBUTIONS

Conceptualization, V.S., A.G.S., and D.A.H.; writing – original draft, V.S., A.G.S., and D.A.H.; writing – review & editing, V.S., P.M., A.A., A.G.S., D.A.H., S.L., Y.T.Y., F.S., M.U.K., J.W.L., and J.B.S.; visualization, P.M. and A.A.; supervision, A.G.S. and D.A.H.; project administration, A.G.S. and D.A.H.; funding acquisition., A.G.S. and D.A.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Samad, N.A., Kim, Y., and Siegel, J.B. (2018). On power denials and lost energy opportunities in downsizing battery packs in hybrid electric vehicles. J. Energy Storage *16*, 187–196. https://doi.org/10.1016/j.est.2018.01.013.
2. National Academies of sciences engineering and medicine (2021). Assessment of technologies for improving light-duty vehicle fuel economy 2025–2035 (The National Academies Press). https://www.nap.edu/read/26092/chapter/1.
3. Dave, A., Mitchell, J., Kandasamy, K., Wang, H., Burke, S., Paria, B., Póczos, B., Whitacre, J., and Viswanathan, V. (2020). Autonomous discovery of battery electrolytes with robotic experimentation and machine learning. Cell Rep. Phys. Sci. *1*, 100264. https://doi.org/10.1016/j.xcrp.2020.100264.
4. Reniers, J.M., Mulder, G., and Howey, D.A. (2021). Unlocking extra value from grid batteries using advanced models. J. Power Sources *487*, 229355. https://doi.org/10.1016/j.jpowsour.2020.229355.
5. Harper, G., Sommerville, R., Kendrick, E., Driscoll, L., Slater, P., Stolkin, R., Walton, A., Christensen, P., Heidrich, O., Lambert, S., et al. (2019). Recycling lithium-ion batteries from electric vehicles. Nature *575*, 75–86. https://doi.org/10.1038/s41586-019-1682-5.
6. Plett, G.L. (2016). Battery management systems, Volume II Equivalent-Circuit Methods (Artech House).
7. Schmalstieg, J., Käbitz, S., Ecker, M., and Sauer, D.U. (2014). A holistic aging model for Li(NiMnCo)O2 based 18650 lithium-ion batteries. J. Power Sources *257*, 325–334. https://doi.org/10.1016/j.jpowsour.2014.02.012.
8. Reniers, J.M., Mulder, G., and Howey, D.A. (2019). Review and performance comparison of mechanical- chemical degradation models for lithium-ion batteries. J. Electrochem. Soc. *166*, A3189–A3200. https://doi.org/10.1149/2.0281914jes.
9. Hu, X., Xu, L., Lin, X., and Pecht, M. (2020). Battery lifetime prognostics. Joule *4*, 310–346. https://doi.org/10.1016/j.joule.2019.11.018.
10. Xiong, R., Li, L., and Tian, J. (2018). Towards a smarter battery management system: A critical review on battery state of health monitoring methods. J. Power Sources *405*, 18–29.
11. Berecibar, M., Gandiaga, I., Villarreal, I., Omar, N., Van Mierlo, J., and Van den Bossche, P. (2016). Critical review of state of health estimation methods of Li-ion batteries for real applications. Renew. Sustain. Energy Rev. *56*, 572–587. https://doi.org/10.1016/j.rser.2015.11.042.
12. Farmann, A., Waag, W., Marongiu, A., and Sauer, D.U. (2015). Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles. J. Power Sources *281*, 114–130. https://doi.org/10.1016/j.jpowsour.2015.01.129.
13. Rezvanizaniani, S.M., Liu, Z., Chen, Y., and Lee, J. (2014). Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility. J. Power Sources *256*, 110–124. https://doi.org/10.1016/j.jpowsour.2014.01.085.
14. Han, X., Lu, L., Zheng, Y., Feng, X., Li, Z., Li, J., and Ouyang, M. (2019). A review on the key issues of the lithium ion battery degradation among the whole life cycle. eTransportation *1*, 100005. https://doi.org/10.1016/j.etran.2019.100005.
15. Dubarry, M., and Baure, G. (2020). Perspective on commercial Li-ion battery testing, best practices for simple and effective protocols. Electronics *9*. https://doi.org/10.3390/electronics9010152.
16. Li, Y., Liu, K., Foley, A.M., Zülke, A., Berecibar, M., Nanini-Maury, E., Van Mierlo, J., and Hoster, H.E. (2019). Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review. Renew. Sustain. Energy Rev. *113*, 109254. https://doi.org/10.1016/j.rser.2019.109254.
17. Aykol, M., Gopal, C.B., Anapolsky, A., Herring, P., van Vlijmen, B., Berliner, M., Bazant, M., Braatz, R., Chueh, W., and Storey, B. (2021). Perspective—combining physics and machine learning to predict battery lifetime. J. Electrochem. Soc. *168*, 030525.
18. Wu, B., Widanage, W.D., Yang, S., and Liu, X. (2020). Battery digital twins: perspectives on the fusion of models, data and artificial intelligence for smart battery management systems. Energy and AI *1*, 100016. https://doi.org/10.1016/j.egyai.2020.100016.

19. Finegan, D.P., Zhu, J., Feng, X., Keyser, M., Ulmefors, M., Li, W., Bazant, M.Z., and Cooper, S.J. (2021). The application of data-driven methods and physics-based learning for improving battery safety. Joule *5*, 316–329. https://doi.org/10.1016/j.joule.2020.11.018.

20. Lee, S., Mohtat, P., Siegel, J.B., Stefanopoulou, A.G., Lee, J.-w., and Lee, T.-k. (2020). Estimation error bound of battery electrode parameters with limited data window. IEEE Trans. Ind. Inf. *16*, 3376–3386.

21. Bloom, I., Cole, B.W., Sohn, J.J., Jones, S.A., Polzin, E.G., Battaglia, V.S., Henriksen, G.L., Motloch, C., Richardson, R., Unkelhaeuser, T., et al. (2001). An accelerated calendar and cycle life study of Li-ion cells. J. Power Sources *101*, 238–247. https://doi.org/10.1016/S0378-7753(01)00783-2.

22. Wang, J., Liu, P., Hicks-Garner, J., Sherman, E., Soukiazian, S., Verbrugge, M., Tataria, H., Musser, J., and Finamore, P. (2011). Cycle-life model for graphite-LiFePO4 cells. J. Power Sources *196*, 3942–3948. https://doi.org/10.1016/j.jpowsour.2010.11.134.

23. Schimpe, M., von Kuepach, M.E., Naumann, M., Hesse, H.C., Smith, K., and Jossen, A. (2018). Comprehensive modeling of temperature-dependent degradation mechanisms in lithium iron phosphate batteries. J. Electrochem. Soc. *165*, A181–A193. https://doi.org/10.1149/2.1181714jes.

24. De Gennaro, M., Paffumi, E., Martini, G., Giallonardo, A., Pedroso, S., and Loiselle-Lapointe, A. (2020). A case study to predict the capacity fade of the battery of electrified vehicles in real-world use conditions. Case Studies on Transport Policy *8*, 517–534. https://doi.org/10.1016/j.cstp.2019.11.005.

25. Ecker, M., Gerschler, J.B., Vogel, J., Käbitz, S., Hust, F., Dechent, P., and Sauer, D.U. (2012). Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data. J. Power Sources *215*, 248–257. https://doi.org/10.1016/j.jpowsour.2012.05.012.

26. Chaoui, H., and Ibe-Ekeocha, C.C. (2017). State of charge and state of health estimation for lithium batteries using recurrent neural networks. IEEE Trans. Veh. Technol. *66*, 8773–8783. https://doi.org/10.1109/TVT.2017.2715333.

27. Chaoui, H., Ibe-Ekeocha, C.C., and Gualous, H. (2017). Aging prediction and state of charge estimation of a LiFePO4 battery using input time-delayed neural networks. Electr. Power Syst. Res. *146*, 189–197. https://doi.org/10.1016/j.epsr.2017.01.032.

28. Ding, Y., Lu, C., and Ma, J. (2017). Li-ion battery health estimation based on multi-layer characteristic fusion and deep learning. IEEE Vehicle Power and Propulsion Conference (VPPC), 1–5. https://doi.org/10.1109/VPPC.2017.8331058.

29. Wu, J., Wang, Y., Zhang, X., and Chen, Z. (2016). A novel state of health estimation method of Li-ion battery using group method of data handling. J. Power Sources *327*, 457–464. https://doi.org/10.1016/j.jpowsour.2016.07.065.

30. Ren, L., Zhao, L., Hong, S., Zhao, S., Wang, H., and Zhang, L. (2018). Remaining useful life prediction for lithium-ion battery: a deep learning approach. IEEE Access *6*, 50587–50598. https://doi.org/10.1109/ACCESS.2018.2858856.

31. Yang, D., Zhang, X., Pan, R., Wang, Y., and Chen, Z. (2018). A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. J. Power Sources *384*, 387–395. https://doi.org/10.1016/j.jpowsour.2018.03.015.

32. Klass, V., Behm, M., and Lindbergh, G. (2014). A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation. J. Power Sources *270*, 262–272. https://doi.org/10.1016/j.jpowsour.2014.07.116.

33. Zhang, Y., Tang, Q., Zhang, Y., Wang, J., Stimming, U., and Lee, A.A. (2020). Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nat. Commun. *11*, 1706. https://doi.org/10.1038/s41467-020-15235-7.

34. Tang, X., Liu, K., Wang, X., Gao, F., MacRo, J., and Widanage, W.D. (2020). Model migration neural network for predicting battery aging trajectories. IEEE Trans. Transp. Electrific. *6*, 363–374. https://doi.org/10.1109/TTE.2020.2979547.

35. Richardson, R.R., Osborne, M.A., and Howey, D.A. (2017). Gaussian process regression for forecasting battery state of health. J. Power Sources *357*, 209–219. https://doi.org/10.1016/j.jpowsour.2017.05.004.

36. Wei, J., Dong, G., and Chen, Z. (2018). Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression. IEEE Trans. Ind. Electron. *65*, 5634–5643. https://doi.org/10.1109/TIE.2017.2782224.

37. Hu, C., Jain, G., Schmidt, C., Strief, C., and Sullivan, M. (2015). Online estimation of lithium-ion battery capacity using sparse Bayesian learning. J. Power Sources *289*, 105–113. https://doi.org/10.1016/j.jpowsour.2015.04.166.

38. Wang, Z., Zeng, S., Guo, J., and Qin, T. (2019). State of health estimation of lithium-ion batteries based on the constant voltage charging curve. Energy *167*, 661–669. https://doi.org/10.1016/j.energy.2018.11.008.

39. Eddahech, A., Briat, O., and Vinassa, J.M. (2014). Determination of lithium-ion battery state-of-health based on constant-voltage charge phase. J. Power Sources *258*, 218–227. https://doi.org/10.1016/j.jpowsour.2014.02.020.

40. Wang, Z., Ma, J., and Zhang, L. (2017). State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the Gaussian process regression. IEEE Access *5*, 21286–21295. https://doi.org/10.1109/ACCESS.2017.2759094.

41. Samad, N.A., Kim, Y., Siegel, J.B., and Stefanopoulou, A.G. (2016). Battery capacity fading estimation using a force-based incremental capacity analysis. J. Electrochem. Soc. *163*, A1584–A1594. https://doi.org/10.1149/2.0511608jes.

42. Yang, N., Liu, T., and Zhong, D. (2016). Online estimation of state-of-health for lithium ion batteries based on charge curves 11th International Conference on Reliability (Maintainability and Safety (ICRMS)), pp. 1–8.

43. Lin, H.-T., Liang, T.-J., and Chen, S.-M. (2013). Estimation of battery state of health using probabilistic neural network. IEEE Trans. Ind. Inf. *9*, 679–685. https://doi.org/10.1109/TII.2012.2222650.

44. Richardson, R.R., Birkl, C.R., Osborne, M.A., and Howey, D.A. (2019). Gaussian process regression for in situ capacity estimation of lithium-ion batteries. IEEE Trans. Ind. Inf. *15*, 127–138. https://doi.org/10.1109/TII.2018.2794997.

45. Richardson, R.R., Osborne, M.A., and Howey, D.A. (2019). Battery health prediction under generalized conditions using a Gaussian process transition model. J. Energy Storage *23*, 320–328. https://doi.org/10.1016/j.est.2019.03.022.

46. Yang, J., Xia, B., Huang, W., Fu, Y., and Mi, C. (2018). Online state-of-health estimation for lithium-ion batteries using constant-voltage charging current analysis. Appl. Energy *212*, 1589–1600. https://doi.org/10.1016/j.apenergy.2018.01.010.

47. Chinomona, B., Chung, C., Chang, L.-K., Su, W.-C., and Tsai, M.-C. (2020). Long short-term memory approach to estimate battery remaining useful life using partial data. IEEE Access *8*, 165419–165431. https://doi.org/10.1109/ACCESS.2020.3022505.

48. Yun, Z., and Qin, W.. Remaining Useful Life Estimation of Lithium-Ion Batteries Based on Optimal Time Series Health Indicator. IEEE Access 8, 55447–55461.

49. Greenbank, S., and Howey, D.A. (2021). Automated feature selection for data-driven models of rapid battery capacity fade and end of life. arXiv, arXiv:2101.04440.

50. Lucu, M., Martinez-Laserna, E., Gandiaga, I., Liu, K., Camblong, H., Widanage, W.D., and Marco, J. (2020). Data-driven nonparametric Li-ion battery ageing model aiming at learning from real operation data - part B: cyclinG operation. J. Energy Storage *30*, 101410. https://doi.org/10.1016/j.est.2020.101410.

51. Lucu, M., Martinez-Laserna, E., Gandiaga, I., Liu, K., Camblong, H., Widanage, W.D., and Marco, J. (2020). Data-driven nonparametric Li-ion battery ageing model aiming at learning from real operation data – part A: storage operation. J. Energy Storage *30*, 101409. https://doi.org/10.1016/j.est.2020.101409.

52. Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy *4*, 383–391. https://doi.org/10.1038/s41560-019-0356-8.

53. Fermín-Cueto, P., McTurk, E., Allerhand, M., Medina-Lopez, E., Anjos, M.F., Sylvester, J., et al. (2020). Identification and machine

learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells. Energy and AI *1*, 100006. https://doi.org/10.1016/j.egyai.2020.100006.

54. Sulzer, V., Mohtat, P., Lee, S., Siegel, J.B., and Stefanopoulou, A.G. (2020). Promise and challenges of a data-driven approach for battery lifetime prognostics. arXiv, arXiv:2010.07460.

55. Bloom, I., Jansen, A.N., Abraham, D.P., Knuth, J., Jones, S.A., Battaglia, V.S., and Henriksen, G.L. (2005). Differential voltage analyses of high-power, lithium-ion cells 1. J. Power Sources *139*, 295–303. https://doi.org/10.1016/j.jpowsour.2004.07.021.

56. Dahn, H.M., Smith, A.J., Burns, J.C., Stevens, D.A., and Dahn, J.R. (2012). User-friendly differential voltage analysis freeware for the analysis of degradation mechanisms in Li-ion batteries. J. Electrochem. Soc. *159*, A1405–A1409. https://doi.org/10.1149/2.013209jes.

57. Wang, J., Purewal, J., Liu, P., Hicks-Garner, J., Soukazian, S., Sherman, E., Sorenson, A., Vu, L., Tataria, H., and Verbrugge, M.W. (2014). Degradation of lithium ion batteries employing graphite negatives and nickel-cobalt-manganese oxide + spinel manganese oxide positives: part 1, aging mechanisms and life estimation. J. Power Sources *269*, 937–948. https://doi.org/10.1016/j.jpowsour.2014.07.030.

58. Lee, S., Siegel, J.B., Stefanopoulou, A.G., Lee, J.-W., and Lee, T.-K. (2020). Electrode state of health estimation for lithium ion batteries considering half-cell potential change due to aging. J. Electrochem. Soc. *167*, 090531. https://doi.org/10.1149/1945-7111/ab8c83.

59. Dubarry, M., Baure, G., and Devie, A. (2018). Durability and reliability of EV batteries under electric utility grid operations: path dependence of battery degradation. J. Electrochem. Soc. *165*, A773–A783. https://doi.org/10.1149/2.0421805jes.

60. Birkl, C.R., Roberts, M.R., McTurk, E., Bruce, P.G., and Howey, D.A. (2017). Degradation diagnostics for lithium ion cells. J. Power Sources *341*, 373–386. https://doi.org/10.1016/j.jpowsour.

61. Zhang, D., Dey, S., Perez, H.E., and Moura, S.J. (2017). Remaining useful life estimation of lithium-ion batteries based on thermal dynamics. 2017 American Control Conference (ACC) pp. 4042–4047, https://doi.org/10.1109/ACCESS.

62. Chu, A., Allam, A., Cordoba Arenas, A., Rizzoni, G., and Onori, S. (2020). Stochastic capacity loss and remaining useful life models for lithium-ion batteries in plug-in hybrid electric vehicles. J. Power Sources *478*, 228991. https://doi.org/10.1016/j.jpowsour.2020.228991.

63. Li, Y., Vilathgamuwa, M., Farrell, T., Choi, S.S., Tran, N.T., and Teague, J. (2019). A physics-based distributed-parameter equivalent circuit model for lithium-ion batteries. Electrochim. Acta *299*, 451–469. https://doi.org/10.1016/j.electacta.2018.12.167.

64. Ouyang, M., Feng, X., Han, X., Lu, L., Li, Z., and He, X. (2016). A dynamic capacity degradation model and its applications considering varying load for a large format Li-ion battery. Appl. Energy *165*, 48–59. https://doi.org/10.1016/j.apenergy.2015.12.063.

65. Atalay, S., Sheikh, M., Mariani, A., Merla, Y., Bower, E., and Widanage, W.D. (2020). Theory of battery ageing in a lithium-ion battery: capacity fade, nonlinear ageing and lifetime prediction. J. Power Sources *478*, 229026. https://doi.org/10.1016/j.jpowsour.2020.229026.

66. Yang, X.G., Leng, Y., Zhang, G., Ge, S., and Wang, C.Y. (2017). Modeling of lithium plating induced aging of lithium-ion batteries: transition from linear to nonlinear aging. J. Power Sources *360*, 28–40. https://doi.org/10.1016/j.jpowsour.2017.05.110.

67. Keil, J., and Jossen, A. (2020). Electrochemical modeling of linear and nonlinear aging of lithium-ion cells. J. Electrochem. Soc. *167*, 110535. https://doi.org/10.1149/1945-7111/aba44f.

68. Ning, G., White, R.E., and Popov, B.N. (2006). A generalized cycle life model of rechargeable Li-ion batteries. Electrochim. Acta *51*, 2012–2022. https://doi.org/10.1016/j.electacta.2005.06.033.

69. Ekström, H., and Lindbergh, G. (2015). A model for predicting capacity fade due to SEI formation in a commercial graphite/LiFePO 4 cell. J. Electrochem. Soc. *162*, A1003–A1007. https://doi.org/10.1149/2.0641506jes.

70. Safari, M., and Delacourt, C. (2011). Simulation-based analysis of aging phenomena in a commercial graphite/LiFePO4 cell. J. Electrochem. Soc. *158*, A1436. https://doi.org/10.1149/2.103112jes.

71. Mohtat, P., Lee, S., Siegel, J.B., and Stefanopoulou, A.G. (2019). Towards better estimability of electrode- specific state of health: decoding the cell expansion. J. Power Sources *427*, 101–111. https://doi.org/10.1016/j.jpowsour.2019.03.104.

72. Hsieh, A.G., Bhadra, S., Hertzberg, B.J., Gjeltema, P.J., Goy, A., Fleischer, J.W., and Steingart, D.A. (2015). Electrochemical-acoustic time of flight: in operando correlation of physical dynamics with battery charge and health. Energy Environ. Sci. *8*, 1569–1577. https://doi.org/10.1039/C5EE00111K.

73. Raj, T., Wang, A.A., Monroe, C.W., and Howey, D.A. (2020). Investigation of path-dependent degradation in lithium-ion batteries. Batteries Supercaps *3*, 1377–1385. https://doi.org/10.1002/batt.202000160.

74. Su, L., Zhang, J., Huang, J., Ge, H., Li, Z., Xie, F., and Liaw, B.Y. (2016). Path dependence of lithium ion cells aging under storage conditions. J. Power Sources *315*, 35–46. https://doi.org/10.1016/j.jpowsour.2016.03.043.

75. Gering, K.L., Sazhin, S.V., Jamison, D.K., Michelbacher, C.J., Liaw, B.Y., Dubarry, M., and andCugnet, M. (2011). Investigation of path dependence in commercial lithium-ion cells chosen for plug-in hybrid vehicle duty cycle protocols. J. Power Sources *196*, 3395–3403. https://doi.org/10.1016/j.jpowsour.2010.05.058.

76. Ma, Z., Jiang, J., Shi, W., Zhang, W., and Mi, C.C. (2015). Investigation of path dependence in commercial lithium-ion cells for pure electric bus applications: aging mechanism identification. J. Power Sources *274*, 29–40. https://doi.org/10.1016/j.jpowsour.2014.10.006.

77. United States Advanced Battery Consortium (2014). Research and development of high-power and high-energy electrochemical storage devices. https://doi.org/10.2172/1160224.

78. Smith, K., Saxon, A., Keyser, M., Lundstrom, B., Cao, Z., and Roc, A. (2017). Life prediction model for grid- connected Li-ion battery energy storage system 2017 American Control Conference (ACC), pp. 4062–4068. https://doi.org/10.23919/ACC.2017.7963578.

79. Li, W., SenGupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D.U. (2021). Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. J. Power Sources *482*, 228863. https://doi.org/10.1016/j.jpowsour.2020.228863.

80. Diao, W., Saxena, S., Han, B., and Pecht, M. (2019). Algorithm to determine the knee point on capacity fade curves of lithium-ion cells. Energies *12*. https://doi.org/10.3390/en12152910.

81. Drummond, R., Zhao, S., Howey, D.A., and Duncan, S.R. (2017). Circuit synthesis of electrochemical supercapacitor models. J. Energy Storage *10*, 48–55. https://doi.org/10.1016/j.est.2016.11.003.

82. Aitio, A., and Howey, D. (2020). Combining non-parametric and parametric models for stable and computationally efficient battery health estimation. ASME 2020 Dynamic Systems and Control Conference. https://doi.org/10.1115/DSCC2020-3180.

83. Marquis, S.G., Sulzer, V., Timms, R., Please, C.P., and Chapman, S.J. (2019). An asymptotic derivation of a single particle model with electrolyte. J. Electrochem. Soc. *166*, A3693–A3706. https://doi.org/10.1149/2.0341915jes.

84. Sulzer, V., Chapman, S.J., Please, C.P., Howey, D.A., and Monroe, C.W. (2019). Faster lead-acid battery simulations from porous-electrode theory: part II. Asymptotic analysis. J. Electrochem. Soc. *166*, A2372–A2382. https://doi.org/10.1149/2.0441908jes.

85. Di Domenico, D., Stefanopoulou, A.G., and Fiengo, G. (2010). Lithium-ion battery state of charge and critical surface charge estimation using an electrochemical model-based extended Kalman filter. J. Dyn. Syst. Meas. Control *132*, 061302. https://doi.org/10.1115/1.4002475.

86. Single, F., Latz, A., and Horstmann, B. (2018). Identifying the mechanism of continued growth of the solid–electrolyte interphase. ChemSusChem *11*, 1950–1955. https://doi.org/10.1002/cssc.201800077.

87. Tang, M., Lu, S., and Newman, J. (2012). Experimental and theoretical investigation of solid-electrolyte- interphase formation mechanisms on glassy carbon. J. Electrochem. Soc. *159*, A1775–A1785. https://doi.org/10.1149/2.025211jes.

88. Ramadass, P., Haran, B., Gomadam, P.M., White, R., and Popov, B.N. (2004). Development of first principles capacity fade model for Li-ion cells. J. Electrochem. Soc. *151*, A196. https://doi.org/10.1149/1.1634273.

89. O'Kane, S.E.J., Campbell, I.D., Marzook, M.W.J., Offer, G.J., and Marinescu, M. Physical origin of the differential voltage minimum associated with lithium plating in Li-ion batteries. J. Electrochem. Soc. 167, 090540, 10.1149/1945-7111/ab90ac.

90. Arora, P., Doyle, M., and White, R.E. (1999). Mathematical modeling of the lithium deposition overcharge reaction in lithium-ion batteries using carbon-based negative electrodes. J. Electrochem. Soc. *146*, 3543–3553. https://doi.org/10.1149/1.1392512.

91. Ai, W., Kraft, L., Sturm, J., Jossen, A., and Wu, B. (2020). Electrochemical thermal-mechanical modelling of stress inhomogeneity in lithium-ion pouch cells. J. Electrochem. Soc. *167*, 013512. https://doi.org/10.1149/2.0122001JES.

92. Chen, C.-H., Brosa Planella, F., O'Regan, K., Gastol, D., Widanage, W.D., and Kendrick, E. Development of experimental techniques for parameterization of multi-scale lithium-ion battery models. J. Electrochem. Soc. 167, 080534, 10.1149/1945-7111/ab9050.

93. Ecker, M., Käbitz, S., Laresgoiti, I., and Sauer, D.U. (2015). Parameterization of a physico-chemical model of a lithium-ion battery. J. Electrochem. Soc. *162*, A1849–A1857. https://doi.org/10.1149/2.0541509jes.

94. Sulzer, V., Marquis, S.G., Timms, R., Robinson, M., and Chapman, S.J. (2021). Python Battery Mathematical Modelling (PyBaMM). J. Open Res. Software *9*, 14. https://doi.org/10.5334/jors.309.

95. Hui, Y., Li, M., Downey, A., Shen, S., Pavan, V., Ye, H., Vanelzen, C., Jain, G., Hu, S., Laflamme, S., and Hu, C. (2021). Physics-based prognostics of implantable-grade lithium-ion battery for remaining usefu life prediction. Journal of Power Sources *485*, 229327. https://doi.org/10.1016/j.jpowsour.2020.229327.

96. An, F., and Ross, M. (1993). A model of fuel economy and driving patterns. SAE Technical Paper 930328. https://doi.org/10.4271/930328.

97. Liaw, B.Y., and Dubarry, M. (2007). From driving cycle analysis to understanding battery performance in real-life electric hybrid vehicle operation. J. Power Sources *174*, 76–88.

98. Saxena, A., Celaya, J.R., Roychoudhury, I., Saha, B., Saha, S., and Goebel, K. (2012). Designing data-driven battery prognostic approaches for variable loading profiles: some lessons learned. First European Conference of the Prognostics and Health Management Society 2012, pp. 10.

99. Song, L., Zhang, K., Liang, T., Han, X., and Zhang, Y. (2020). Intelligent state of health estimation for lithium- ion battery pack based on big data analysis. J. Energy Storage *32*, 101836. https://doi.org/10.1016/j.est.2020.101836.

100. Wang, Q., Wang, Z., Zhang, L., Liu, P., and Zhang, Z. (2020). A novel consistency evaluation method for series-connected battery systems based on real-world operation data. IEEE Trans. Transp. Electrific. *7*, 437–451. https://doi.org/10.1109/TTE.2020.3018143.

101. She, C., Wang, Z., Sun, F., Liu, P., and Zhang, L. (2020). Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network. IEEE Trans. Ind. Inf. *16*, 3345–3354. https://doi.org/10.1109/TII.2019.2951843.

102. Huo, Q., Ma, Z., Zhao, X., Zhang, T., and Zhang, Y. (2021). Bayesian network based state-of-health estimation for battery on electric vehicle application and its validation through real-world data. IEEE Access *9*, 11328–11341. https://doi.org/10.1109/ACCESS.2021.3050557.

103. De Angelis, V., Preger, Y., and Chalama, B.R. (2021). Battery lifecycle framework: a flexible repository and visualization tool for battery data from materials development to field implementation. ECSarXiv. https://doi.org/10.1149/osf.io/h7c24.

104. Chao, M.A., Kulkarni, C., Goebel, K., and Fink, O. (2020). Fusing physics-based and deep learning models for prognostics. arXiv, arXiv:2003.00732.

105. Parish, E.J., and Duraisamy, K. (2016). A paradigm for data-driven predictive modeling using field inversion and machine learning. J. Comp. Phys. *305*, 758–774. https://doi.org/10.1016/j.jcp.2015.11.012.

106. Raissi, M., Perdikaris, P., and Karniadakis, G.E. (2017). Physics informed deep learning (part II) data-driven discovery of nonlinear partial differential equations. arXiv, arXiv:1711.10566.

107. Raissi, M., Perdikaris, P., and Karniadakis, G.E. (2017). Physics informed deep learning (Part I): data-driven solutions of nonlinear partial differential equations. arXiv, arXiv:1711.10561.

108. Raissi, M., Perdikaris, P., and Karniadakis, G.E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comp. Phys. *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

109. Chen, R.T.Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. (2018). Neural ordinary differential equations 32nd Conference on Neural Information Processing Systems. https://doi.org/10.1007/978-3-030-15679-4.

110. Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., and Ramadhan, A. (2020). Universal differential equations for scientific machine learning. arXiv, arXiv:2001.04385.

111. Rackauckas, C., Innes, M., Ma, Y., Bettencourt, J., White, L., and Dixit, V. (2019). DiffEqFlux.jl — a Julia library for neural differential equations. arXiv, arXiv:1902.02376.

112. Dandekar, R., Rackauckas, C., and Barbastathis, G. (2020). A machine learning-aided global diagnostic and comparative tool to assess effect of quarantine control in COVID-19 spread. Patterns (N Y) *1*, 100145. https://doi.org/10.1016/j.patter.2020.100145.

113. Bills, A., Sripad, S., Fredericks, W.L., Guttenberg, M., Charles, D., Frank, E., and Viswanathan, V. (2020). Universal battery performance and degradation model for electric aircraft. arXiv. https://doi.org/10.26434/chemrxiv.12616169.

114. Tait, D.J., Brosa Planella, F., Damoulas, T., and Dhammika Widanage, W. (2020). Scalable multitask latent force models with applications to predicting lithium-ion concentration. NeurIPS.

115. Northvolt. http://web.archive.org/web/20210630190549/https://northvolt.com/articles/northvolt-fluence-may2021/.

116. Speltino, C., Stefanopoulou, A., and Fiengo, G. (2010). Cell equalization in battery stacks through State of Charge estimation polling. Proceedings of the 2010 American Control Conference, pp. 5050–5055, https://doi.org/10.1109/acc.2010.5530710.

117. Song, Z., Hou, J., Li, X., Wu, X., Hu, X., Hofmann, H., and Sun, J. (2020). The sequential algorithm for combined state of charge and state of health estimation of lithium-ion battery based on active current injection. Energy *193*, 116732. https://doi.org/10.1016/j.energy.2019.116732.

118. Xiong, R., Cao, J., Yu, Q., He, H., and Sun, F. (2017). Critical review on the battery state of charge estimation methods for electric vehicles. IEEE Access *6*, 1832–1843. https://doi.org/10.1109/ACCESS.2017.2780258.

119. Blei, D.M., Kucukelbir, A., and McAuliffe, J.D. (2017). Variational inference: a review for statisticians. J. Am. Stat. Assoc. *112*, 859–877. https://doi.org/10.1080/01621459.2017.1285773.

120. Craiu, R.V., and Rosenthal, J.S. (2014). Bayesian computation via markov chain monte carlo. Annual Review of Statistics and Its Application *1*, 179–201. https://doi.org/10.1146/annurev-statistics-022513-115540.

121. Salvatier, J., Wiecki, T.V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. PeerJ Comput. Sci. *2*, e55.

122. Martinez-Laserna, E., Gandiaga, I., Sarasketa-Zabala, E., Badeda, J., Stroe, D.-I., Swierczynski, M., and Goikoetxea, A. (2018). Battery second life: hype, hope or reality? A critical review of the state of the art. Renew. Sustain. Energy Rev. *93*, 701–718. https://doi.org/10.1016/j.rser.2018.04.035.

123. European Commission (2020). Green deal: sustainable batteries for a circular and climate neutral economy. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2312.

124. Martinez-Laserna, E., Sarasketa-Zabala, E., Villarreal Sarria, I., Stroe, D.I., Swierczynski, M., Warnecke, A., Timmermans, J.M., Goutam, S., Omar, N., and Rodriguez, P. (2018). Technical viability of battery second life: a study from the ageing perspective. IEEE Transactions on Industry Applications *54*, 2703–2713. https://doi.org/10.1109/TIA.2018.2801262.

125. Park, S., Kato, D., Gima, Z., Klein, R., and Moura, S. (2018). Optimal experimental design for parameterization of an electrochemical lithium-ion battery model. J. Electrochem. Soc. *165*, A1309–A1323. https://doi.org/10.1149/2.0421807jes.

126. Bizeray, A.M., Kim, J.H., Duncan, S.R., and Howey, D.A. (2019). Identifiability and parameter estimation of the single particle lithium-ion battery model. IEEE Trans. Contr. Syst. Technol. *27*, 1862–1877.

127. Zhang, C., Wang, Y., Gao, Y., Wang, F., Mu, B., and Zhang, W. (2019). Accelerated fading recognition for lithium-ion batteries with nickel-cobalt-manganese cathode using quantile regression method. Appl. Energy *256*, 113841. https://doi.org/10.1016/j.apenergy.2019.113841.

128. Cui, X., Siegel, J., Stefanopoulou, A., and Avestruz, A.-T. (2021). Grid interfaces to electric vehicle chargers using statistically-structured power conversion architectures for second-use batteries as energy buffering. arXiv, arXiv:2104.14976.

129. Neubauer, J., Smith, K., Wood, E., and Pesaran, A. (2015). Identifying and overcoming critical barriers to widespread second use of PEV batteries. Natl. Renew. Energy Lab. (NREL) *2*, 23–62. https://doi.org/10.2172/1171780.

130. Opitz, D., and Maclin, R. (1999). Popular ensemble methods: an empirical study. J. Artif. Intell. Res. *11*, 169–198. https://doi.org/10.1613/jair.614.

131. Baker, J.A., Beuse, M., Decaluwe, S.C., Jing, L.W., Khoo, E., Sripad, S., Ulissi, U., Verma, A., Wang, A.A., Yeh, Y.T., et al. (2020). Fostering a sustainable community in batteries. ACS Energy Lett *5*, 2361–2366. https://doi.org/10.1021/acsenergylett.0c01304.

132. Frigola, R., Chen, Y., and Rasmussen, C.E. (2014). Variational Gaussian process state-space models. Advances in Neural Information Processing Systems *4*, 3680–3688.

133. Howey, D.A., Roberts, S.A., Viswanathan, V., Mistry, A., Beuse, M., Khoo, E., Decaluwe, S.C., and andSulzer, V. (2020). Free radicals: making a case for battery modeling. Electrochem. Soc. Interface *29*, 30.